

# Mapping the Region of Entropic Vectors with Support Enumeration & Information Geometry

Yunshu Liu   John MacLaren Walsh

Dept. of ECE, Drexel University, Philadelphia, PA 19104, USA

yunshu.liu@drexel.edu   jwalsh@coe.drexel.edu

## Abstract

The region of entropic vectors is a convex cone that has been shown to be at the core of many fundamental limits for problems in multiterminal data compression, network coding, and multimedia transmission. This cone has been shown to be non-polyhedral for four or more random variables, however its boundary remains unknown for four or more discrete random variables. Methods for specifying probability distributions that are in faces and on the boundary of the convex cone are derived, then utilized to map optimized inner bounds to the unknown part of the entropy region. The first method utilizes tools and algorithms from abstract algebra to efficiently determine those supports for the joint probability mass functions for four or more random variables that can, for some appropriate set of non-zero probabilities, yield entropic vectors in the gap between the best known inner and outer bounds. These supports are utilized, together with numerical optimization over non-zero probabilities, to provide inner bounds to the unknown part of the entropy region. Next, information geometry is utilized to parameterize and study the structure of probability distributions on these supports yielding entropic vectors in the faces of entropy and in the unknown part of the entropy region.

## I. INTRODUCTION

The region of entropic vectors is a convex cone that is known to be at the core of many yet undetermined fundamental limits for problems in data compression, network coding, and multimedia transmission. This set has been shown to be non-polyhedral, but its boundaries remain unknown. This paper studies methods of parameterizing probability distributions that yield entropic vectors in this unknown part of the entropy region.

In §II, after reviewing the definition of this set and its importance in applications, we discuss the best known outer and inner bounds for it, most of the latter of which are based on results for representable matroids and linear polymatroids. These inner bounds, which, as we review, are based on linear constructions, are unable to parameterize the boundary of the region of entropic vectors because non-linear dependence structures are necessary to parameterize the unknown part of it.

The authors thank the National Science Foundation for its support in part under CCF-1053702. Preliminary versions of this work were presented at the 2013 Information Theory Workshop and Allerton 2015.

Bearing this in mind, in §III, we provide a systematic and efficient method for searching for supports for joint probability mass functions which can, for appropriate choices of non-zero probabilities, yield entropic vectors in between the best known outer and inner bounds. Key to this study is determining equivalence classes of supports, within which each of the contained supports are guaranteed, when ranging over all non-zero probabilities, to give the same sets of entropy vectors. Only one representative support from each of these equivalence classes, the canonical representative, is necessary to be considered for the purposes of mapping entropies. These equivalence classes are formalized via the language of group actions, and an efficient algorithm for listing canonical supports is provided.

With these canonical supports in hand, §IV sets about determining which of the supports for four random variables yield entropic vectors in the gap between the best known inner and outer bounds on the entropy region. Via numerical optimization over the probabilities on these supports, tuned inner bounds to unknown part of the entropy region are also provided.

Seeking to gain a better understanding of the structure of probability distributions that would yield extremal entropic vectors in this unknown part of the entropy region, we shift in §V to studying the information geometric properties of probability distributions associated with entropic vectors that are extremal in the sense that they live in faces of the widely used outer bound. Furthermore, those distributions on the smallest support that can yield entropic vectors in the unknown part of the entropy region are shown to have a special information geometric structure, which we also show disappears for larger supports having this property. The paper concludes in §VI with a number of interesting directions for further investigation.

## II. BOUNDING THE REGION OF ENTROPIC VECTORS

Consider  $N$  discrete random variables  $\mathbf{X} = \mathbf{X}_{\mathcal{N}} = (X_1, \dots, X_N)$ ,  $\mathcal{N} = \{1, \dots, N\}$  with joint probability mass function  $p_{\mathbf{X}}(\mathbf{x})$ . To every non-empty subset of these random variables  $\mathbf{X}_{\mathcal{A}} := (X_n \mid n \in \mathcal{A})$ ,  $\mathcal{A} \subseteq \mathcal{N}$ , there is associated a Shannon entropy  $H(\mathbf{X}_{\mathcal{A}})$  calculated from the marginal distribution  $p_{\mathbf{X}_{\mathcal{A}}}(\mathbf{x}_{\mathcal{A}}) = \sum_{\mathbf{x}_{\mathcal{N} \setminus \mathcal{A}}} p_{\mathbf{X}_{\mathcal{N}}}(\mathbf{x})$  via

$$H(\mathbf{X}_{\mathcal{A}}) = \sum_{\mathbf{x}_{\mathcal{A}}} -p_{\mathbf{X}_{\mathcal{A}}}(\mathbf{x}_{\mathcal{A}}) \log_2 p_{\mathbf{X}_{\mathcal{A}}}(\mathbf{x}_{\mathcal{A}}) \quad (1)$$

One can stack these entropies of different non-empty subsets into a  $2^N - 1$  dimensional vector  $\mathbf{h} = (H(\mathbf{X}_{\mathcal{A}}))_{\mathcal{A} \subseteq \mathcal{N}}$ , which can be clearly viewed as  $\mathbf{h}(p_{\mathbf{X}})$ , a function of the joint distribution  $p_{\mathbf{X}}$ . A vector  $\mathbf{h}_? \in \mathbb{R}^{2^N - 1}$  is said to be *entropic* if there is some joint distribution  $p_{\mathbf{X}}$  such that  $\mathbf{h}_? = \mathbf{h}(p_{\mathbf{X}})$ . The region of entropic vectors is then the image of the set  $\mathcal{D} = \{p_{\mathbf{X}} \mid p_{\mathbf{X}}(\mathbf{x}) \geq 0, \sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) = 1\}$  of valid joint probability mass functions under the function  $\mathbf{h}(\cdot)$ , and is denoted by

$$\Gamma_N^* = \mathbf{h}(\mathcal{D}) \subsetneq \mathbb{R}^{2^N - 1} \quad (2)$$

It is known that the closure of this set  $\bar{\Gamma}_N^*$  is a convex cone [1], but surprisingly little else is known about this cone for  $N \geq 4$ . Understanding the “shape” and boundaries of the set  $\bar{\Gamma}_N^*$  is the subject of this paper.

The fundamental importance of  $\bar{\Gamma}_N^*$  lies in several contexts in signal processing, compression, network coding and information theory [1]. Firstly, the fundamental rate region limits of many data compression problems that otherwise remain unsolved can be directly expressed in terms of  $\bar{\Gamma}_N^*$  and related entropy vector sets. In particular, a class of problems studied by Zhang and Yeung under the name of distributed source coding for satellite communications [1], [2] have rate regions directly expressible as projections of  $\bar{\Gamma}_N^*$  after intersection with linear entropy equalities associated with certain Markov chain conditions. Secondly, the multiple multicast capacity region for a lossless network under network coding can also be directly expressed in terms of  $\bar{\Gamma}_N^*$ , as was proved in [3] and [1], [4]. The multi-source network coding (MSNC) problem generalizes the YZDSC problem by allowing for an arbitrary topology of intermediate encoding nodes between the source encoders and decoders. If one can determine the boundaries of  $\bar{\Gamma}_N^*$  one can determine the capacity region of any network under network coding. Furthermore, Chan and Grant [5]–[7] proved that every part of the (unknown) boundary of  $\bar{\Gamma}_N^*$  has a network whose capacity region depends on it. Hence, they have established that the problem of determining the boundaries of  $\bar{\Gamma}_N^*$  is equivalent to the problem of calculating the capacity region of every network under network coding.

Even more generally, as *all* achievable rate regions in information theory are expressed in terms of information measures between random variables obeying certain distribution constraints, they can be expressed as linear projections of entropy vectors associated with these *constrained* random variables. In the case of network coding and the distributed source coding problems introduced in the previous two sections, these constraints are embodied solely as entropy constraints and conditional entropies being zero, and hence can handle, after a great deal of proof, the constraints solely in entropy space and work with the region of unconstrained entropic vectors. However, as one moves to more general multi-terminal information theory problems, particular distributions, either marginal or conditional, are typically specified, and in many cases there are also distortion/expectation of some function of a collection of the variables constraints. Such constraints which can only be expressed in probability distribution space make the problem even harder, and a modified region of entropic vectors  $\bar{\Gamma}_N^*(\mathcal{C})$  will need to be employed incorporating distribution constraints  $\mathcal{C}$ . However, given the wide array of multiterminal information theory whose rate regions can be directly expressed in terms of the simpler unconstrained region  $\bar{\Gamma}_N^*$ , it makes sense to attempt to bound and characterize it first. Thus, the remaining sections of this paper will attempt to review what is known about bounding  $\bar{\Gamma}_N^*$ , as well as introduce some new relevant ideas from abstract algebra, combinatorics, and information geometry to understand these bounds, and enabling the creation of new even better bounds.

#### A. Outer Bounds of $\bar{\Gamma}_N^*$

Viewed as a function  $h_{\mathcal{A}} = H(\mathbf{X}_{\mathcal{A}})$  of the selected subset, with the convention that  $h_{\emptyset} = 0$ , entropy is *sub-modular* [1], [8], meaning that

$$h_{\mathcal{A}} + h_{\mathcal{B}} \geq h_{\mathcal{A} \cap \mathcal{B}} + h_{\mathcal{A} \cup \mathcal{B}} \quad \forall \mathcal{A}, \mathcal{B} \subseteq \mathcal{N}, \quad (3)$$

and is also *non-decreasing* and *non-negative*, meaning that

$$h_{\mathcal{K}} \geq h_{\mathcal{W}} \geq 0 \quad \forall \mathcal{W} \subseteq \mathcal{K} \subseteq \mathcal{N}. \quad (4)$$

Viewed as requirements for arbitrary set functions (not necessarily entropy) the inequalities (3) and (4) are known as the *polymatroidal axioms* [1], [8], and a function obeying them is called the *rank function* of a *polymatroid*. If a set function  $f$  that obeys the polymatroidal axioms (3) and (4) additionally obeys

$$f_{\mathcal{A}} \leq |\mathcal{A}|, \quad f_{\mathcal{A}} \in \mathbb{Z} \quad \forall \mathcal{A} \subseteq \mathcal{N} \quad (5)$$

then it is the rank function of a *matroid* on the ground set  $\mathcal{N}$ .

Since entropy must obey the polymatroidal axioms, the set of all rank functions of polymatroids forms an outer bound for the region of entropic vectors which is often denoted by

$$\Gamma_N = \left\{ \mathbf{h} \left| \begin{array}{l} \mathbf{h} \in \mathbb{R}^{2^N - 1} \\ h_{\mathcal{A}} + h_{\mathcal{B}} \geq h_{\mathcal{A} \cap \mathcal{B}} + h_{\mathcal{A} \cup \mathcal{B}} \quad \forall \mathcal{A}, \mathcal{B} \subseteq \mathcal{N} \\ h_{\mathcal{K}} \geq h_{\mathcal{W}} \geq 0 \quad \forall \mathcal{W} \subseteq \mathcal{K} \subseteq \mathcal{N} \end{array} \right. \right\} \quad (6)$$

In fact, any inequality in (6) can be expressed as a sum of the following two types of elemental inequalities [1]

$$\begin{aligned} h_{\mathcal{N}} - h_{\mathcal{N} \setminus i} &\geq 0, i \in \mathcal{N} \\ h_{i\mathcal{K}} + h_{j\mathcal{K}} - h_{\mathcal{K}} - h_{ij\mathcal{K}} &\geq 0, \text{ for } i \neq j, \mathcal{K} \subset \mathcal{N} \setminus ij \end{aligned} \quad (7)$$

The inequalities (7) are the minimal, non-redundant, set of information inequalities for defining  $\Gamma_N$ . As we can see from the definition,  $\Gamma_N$  is a polyhedron, and this polyhedral set is often known as the Shannon outer bound for  $\bar{\Gamma}_N^*$  [1], [8].

While in the low dimensional cases we have  $\Gamma_2 = \Gamma_2^*$  and  $\Gamma_3 = \bar{\Gamma}_3^*$ , for  $N \geq 4$ ,  $\Gamma_N \neq \bar{\Gamma}_N^*$ . Zhang and Yeung first showed this in [8] by proving a new inequality among 4 variables

$$2I(C; D) \leq I(A; B) + I(A; C, D) + 3I(C; D|A) + I(C; D|B) \quad (8)$$

which held for entropies and was not implied by the polymatroidal axioms, which they dubbed a *non-Shannon type* inequality to distinguish it from inequalities implied by  $\Gamma_N$ . For roughly the next decade a few authors produced other new non-Shannon inequalities [9], [10]. In 2007, Matúš [11] showed that  $\bar{\Gamma}_N^*$  is not a polyhedron for  $N \geq 4$ . The proof of this fact was carried out by constructing a sequence of non-Shannon inequalities, including

$$\begin{aligned} &s[I(A; B|C) + I(A; B|D) + I(C; D) - I(A; B)] \\ &+ I(B; C|A) + \frac{s(s+1)}{2}[I(A; C|B) + I(A; B|C)] \geq 0 \end{aligned} \quad (9)$$

Notice (9) is the same as Zhang-Yeung inequality (8) when  $s = 1$ . Additionally, the infinite sequence of inequalities was used with a curve constructed from a particular form of distributions to prove  $\bar{\Gamma}_N^*$  is not a polyhedron. Despite this result, even  $\bar{\Gamma}_4^*$  is still not fully understood. Since then, many authors has been investigating the properties of  $\bar{\Gamma}_N^*$  with the hope of ultimately fully characterizing the region [12]–[17].

### B. Inner Bounds of $\bar{\Gamma}_N^*$

Shifting from outer bounds to bounding from the inside, the most common way to generate inner bounds for the region of entropic vectors is to consider special families of distributions for which the entropy function is known to have certain properties. [15], [18]–[20] focus on calculating inner bounds based on special properties of binary random variables. However, the most common way to generate inner bounds is based on inequalities for representable matroids [21], boolean polymatroids [22], [23] and subspace arrangements.

For the latter method, we first introduce some basics in linear polymatroids and the Ingleton inner bound. Fix a  $N' \geq N$ , and partition the set  $\{1, \dots, N'\}$  into  $N$  disjoint sets  $\mathcal{Q}_1, \dots, \mathcal{Q}_N$ . Let  $\mathbf{U}$  be a length  $m$  row vector whose elements are i.i.d. uniform over the finite field  $GF(q)$ , and let  $\mathbf{T}$  be a particular  $m \times N'$  deterministic matrix with elements in  $GF(q)$ . Consider the  $N'$  dimensional vector

$$\mathbf{Y} = \mathbf{U}\mathbf{T}, \text{ and define } \mathbf{X}_i = \mathbf{Y}_{\mathcal{Q}_i}, \quad i \in \{1, \dots, N\}.$$

The subset entropies of the random variables  $\{\mathbf{X}_i\}$  obey

$$H(\mathbf{X}_{\mathcal{A}}) = g(\mathcal{A}) \log_2(q) = \text{rank}([\mathbf{T}_{\mathcal{Q}_i} | i \in \mathcal{A}]) \log_2(q). \quad (10)$$

A set function  $g(\cdot)$  created in such a manner is called a linear polymatroid or a subspace rank function. It obeys the polymatroidal axioms, and is additionally proportional to an integer valued vector. If the  $\mathcal{Q}_i$  are all singletons and  $N' = N$ , then this set function is proportional (via  $\log q$ ) to the rank function of a representable matroid [24]. Alternatively, when the sets  $\mathcal{Q}_i$  are not singletons and  $N' > N$ , such a construction is clearly related to a representable matroid on a larger ground set [24]. Indeed, the subspace rank function vector is merely formed by taking some of the elements from the  $2^{N'} - 1$  representable matroid rank function vector associated with  $\mathbf{T}$ . That is, set function vectors created via (10) are ( $\log_2 q$  scaled) projections of rank function vectors of representable matroids.

Set functions capable of being represented in this manner for some  $N', q$  and  $\mathbf{T}$ , are called subspace ranks in some contexts [25]–[27], while other papers effectively define a collection of vector random variables created in this manner a subspace arrangement [28].

Define  $\mathcal{J}_N$  to be the conic hull of all subspace ranks for  $N$  subspaces. It is known that  $\mathcal{J}_N$  is an inner bound for  $\bar{\Gamma}_N^*$  [25], which we name the subspace inner bound. So far  $\mathcal{J}_N$  is only known for  $N \leq 5$  [27], [28]. More specifically,  $\mathcal{J}_2 = \bar{\Gamma}_2^* = \Gamma_2$ ,  $\mathcal{J}_3 = \bar{\Gamma}_3^* = \Gamma_3$ . As with most entropy vector sets, things start to get interesting at  $N = 4$  variables (subspaces). For  $N = 4$ ,  $\mathcal{J}_4$  is given by the Shannon type inequalities (i.e. the polymatroidal axioms) together with six additional inequalities known as *Ingleton's inequality* [25], [26], [29] which states that for  $N = 4$  random variables

$$Ingleton_{ij} \geq 0 \quad (11)$$

where

$$\begin{aligned} \text{Ingleton}_{ij} &= I(X_k; X_l | X_i) + I(X_k; X_l | X_j) \\ &\quad + I(X_i; X_j) - I(X_k; X_l) \end{aligned}$$

Thus,  $\mathcal{I}_4$  is usually called the Ingleton inner bound.

### C. Semimatroids and Faces of $\Gamma_N$ and $\Gamma_N^*$

In [26], the concept of *semimatroid* is introduced to help analyze the conditional independences among four random variables. In this section, we first review the definition of a *semimatroid*; then some results in [26] that are highly related to the structure of the region of entropic vectors on four variables will be presented; at last, we build the mapping between subset of extreme rays of  $\Gamma_4$  and some particular *p-representable* semimatroids, which we will use information geometry to analyze in section §V.

Let  $\mathcal{N} = \{1, 2, \dots, N\}$  and  $\mathcal{S}$  be the family of all couples  $(i, j | \mathcal{K})$ , where  $\mathcal{K} \subset \mathcal{N}$  and  $ij$  is the union of two singletons  $i$  and  $j$  in  $\mathcal{N} \setminus \mathcal{K}$ . If we include the cases when  $i = j$ , there are, for example, 18 such couples for three variables, and 56 such couples for  $N = 4$ . A relation  $\mathcal{L} \subset \mathcal{S}$  is called *probabilistically representable* or *p-representable* if there exists a system of  $N$  random variables  $\mathbf{X} = \{X_i\}_{i \in \mathcal{N}}$  such that

$$\begin{aligned} \mathcal{L} &= \{(i, j | \mathcal{K}) \in \mathcal{S}(N) | X_i \text{ is conditionally} \\ &\quad \text{independent of } X_j \text{ given } X_{\mathcal{K}} \text{ i.e. } I(X_i; X_j | X_{\mathcal{K}}) = 0\}. \end{aligned}$$

**Definition 1:** For  $\mathbf{f} \in \Gamma_N$  we define  $||\mathbf{f}||$  as

$$||\mathbf{f}|| = \{(i, j | \mathcal{K}) \in \mathcal{S}(N) | f_{i\mathcal{K}} + f_{j\mathcal{K}} - f_{ij\mathcal{K}} - f_{\mathcal{K}} = 0\}. \quad (12)$$

A relation  $\mathcal{L} \subset \mathcal{S}(N)$  is called a *semimatroid* if and only if  $\mathcal{L} = ||\mathbf{f}||$  for some  $\mathbf{f} \in \Gamma_N$ , the Shannon outer bound for  $N$  random variables.

We use  $\text{Semi}(N)$  to denote the set of all semimatroids on  $N$ , and we say that semimatroid  $\mathcal{L}$ ,  $\mathcal{L} = ||\mathbf{f}||$ , arises from polymatroid vector  $\mathbf{f}$ . The *p-representable* semimatroids are just those semimatroids arising from an entropic vector  $\mathbf{h}$ . We use  $P_{\text{rep}}(N)$  to denote the set of all *p-representable* relations on  $\mathcal{N}$ . For  $N \leq 3$ , since  $\Gamma_N = \bar{\Gamma}_N^*$ , we have  $P_{\text{rep}}(N) = \text{Semi}(N)$ . However,  $P_{\text{rep}}(4) \subsetneq \text{Semi}(4)$ , that is to say, there are semimatroids on four variables that are not *p-representable*. The main theorem of [30] lists all irreducible *p-representable* semimatroids over four variables. There are 120 such semimatroids of 16 types, and every *p-representable* semimatroid is at the intersection of some of these semimatroids. For  $\mathcal{N} = \{1, 2, 3, 4\}$ , with  $\mathcal{K} \subseteq \mathcal{N}$  and  $0 \leq t \leq |\mathcal{N} \setminus \mathcal{K}|$ , define  $\mathbf{r}_t^{\mathcal{K}}$ ,  $\mathbf{g}_i^{(2)}$  and  $\mathbf{g}_i^{(3)}$  such that  $\mathbf{r}_t^{\mathcal{K}}(\mathcal{W})$ ,  $\mathbf{g}_i^{(2)}(\mathcal{W})$  and  $\mathbf{g}_i^{(3)}(\mathcal{W})$  as follows:

$$\mathbf{r}_t^{\mathcal{K}}(\mathcal{W}) = \min\{t, |\mathcal{W} \setminus \mathcal{K}|\} \text{ with } \mathcal{W} \subseteq \mathcal{N}$$

$$g_i^{(2)}(\mathcal{W}) = \begin{cases} 2 & \text{if } \mathcal{W} = i \\ \min\{2, |\mathcal{W}|\} & \text{if } \mathcal{W} \neq i \end{cases}$$

$$g_i^{(3)}(\mathcal{W}) = \begin{cases} |\mathcal{W}| & \text{if } i \notin \mathcal{W} \\ \min\{3, |\mathcal{W}| + 1\} & \text{if } i \in \mathcal{W} \end{cases}$$

Now we present the main theorem of [30]

**Theorem 1:** (Matúš) [30] There are 120 irreducible  $p$ -representable semimatroids of sixteen types over four-element set  $N$ . Among which there are 36 ingletton semimatroids of 11 types:  $[[0]]$ ,  $[[r_1^{N-i}]]$  for  $i \in \mathcal{N}$ ,  $[[r_1^{ij}]]$  for  $i, j \in \mathcal{N}$  distinct,  $[[r_1^i]]$  for  $i \in \mathcal{N}$ ,  $[[r_2^i]]$  for  $i \in \mathcal{N}$ ,  $[[r_2^{ij}]]$  for  $i, j \in \mathcal{N}$  distinct,  $[[r_2]]$ ,  $[[r_3]]$ ,  $[[g_i^{(2)}]]$  for  $i \in \mathcal{N}$ ,  $[[g_i^{(3)}]]$  for  $i \in \mathcal{N}$ . There are also 84 non-Ingletton semimatroids of 5 types:

$$\begin{aligned} \mathcal{L}_{ij}^{kl|\emptyset} &= \{(kl|i), (kl|j), (ij|\emptyset), (kl|ij)\} \\ &\cup \{(k|ij), (l|ij), (ij|kl), (j|ikl), (k|ijl), (l|ijk)\} \\ \mathcal{L}_{ij}^{(ij|kl)} &= \{(ij|k), (ij|l), (kl|ij), (kl|i), (kl|j)\} \\ \mathcal{L}_{ij}^{(ik|jl)} &= \{(kl|ij), (ij|k), (ik|l), (kl|j), (l|ij), (l|ijk)\} \\ \mathcal{L}_{ij}^{ik|j} &= \{(ij|k), (ik|l), (kl|j), (ij|kl), (j|ikl)\} \\ &\cup \{(k|ijl), (l|ijk)\} \\ \mathcal{L}_{ij}^{jl|\emptyset} &= \{(kl|i), (jl|k), (ij|\emptyset), (kl|ij)\} \\ &\cup \{(k|ij), (l|ij), (ij|kl), (j|ikl), (k|ijl), (l|ijk)\} \end{aligned}$$

The theorem not only solved  $p$ -representability of semimatroids over four variables, but also answered the question of which faces of  $\Gamma_4$  have interior points that are entropic, which is stated in the following corollary:

**Corollary 1:** A relation  $\mathcal{L}_0 \subseteq \mathcal{S}(N)$  is a  $p$ -representable semimatroid if and only if

$$\mathcal{F}_0 = \{ \mathbf{h} \in \Gamma_N \mid h_{i\mathcal{K}} + h_{j\mathcal{K}} - h_{\mathcal{K}} - h_{ij\mathcal{K}} = 0, \forall (i, j|\mathcal{K}) \in \mathcal{L} \} \quad (13)$$

is a face of  $\Gamma_N$  such that there exists  $\mathbf{h}_0$ , a point in the relative interior of  $\mathcal{F}_*$ , satisfying  $\mathbf{h}_0 \in \Gamma_N^*$ .

**Proof:**  $\Rightarrow$  Suppose relation  $\mathcal{L}_0$  is a  $p$ -representable semimatroid. By definition (6) and (7),

$$\mathcal{F}_a = \{ \mathbf{h} \in \Gamma_N \mid h_{i\mathcal{K}} + h_{j\mathcal{K}} - h_{\mathcal{K}} - h_{ij\mathcal{K}} = 0 \} \quad (14)$$

defines a facet of  $\Gamma_N$  if  $i \neq j$  or  $i = j$  and  $K = \mathcal{N} \setminus \{i, j\}$ , and an intersection of such facets otherwise. By definition a semimatroid  $\mathcal{L}_0$  must be a union of  $(i, j|\mathcal{K})$  couples such that there is a polymatroid  $\mathbf{f} \in \Gamma_N$  which obeys exclusively these independence relations. Thus  $\mathcal{F}_0$  defined by any semimatroid must be face of  $\Gamma_N$  because it is an exhaustive list of facets whose intersection forms this face. Furthermore, since  $\mathcal{L}_0$  is  $p$ -representable, there exist some collection of random variables, generating an entropic vector in  $\Gamma_n^* \cap \mathcal{F}_0$ , that does not obey any additional

conditional independence relations beyond  $\mathcal{L}_0$ . This vector is thus in the relative interior of  $\mathcal{F}_0$  (for being on a relative boundary of this face would require living in additional facets of  $\Gamma_N$  and thus obeying additional conditional independence relations).

$\Leftarrow$  Now suppose for a given  $\mathcal{F}_0$ , which is a face of  $\Gamma_N$ , we have  $\mathbf{h}_0$ , a relative interior point of  $\mathcal{F}_0$  such that  $\mathbf{h}_0 \in \Gamma_N^*$ . Then the relation corresponding to the information equalities satisfied by  $\mathbf{h}_0$  must be  $p$ -representable.  $\square$

#### D. The gap between $\mathcal{I}_4$ and $\Gamma_4$

For the structure of the gap between  $\mathcal{I}_4$  and  $\Gamma_4$ , we know  $\Gamma_4$  is generated by 28 elemental Shannon type information inequalities [1]. As for  $\mathcal{I}_4$ , in addition to the 28 Shannon type information inequalities, we also need six Ingleton's inequalities (11), thus  $\mathcal{I}_4 \subsetneq \Gamma_4$ . In [26] it is stated that  $\Gamma_4$  is the disjoint union of  $\mathcal{I}_4$  and six cones  $\{\mathbf{h} \in \Gamma_4 | \text{Ingleton}_{ij} < 0\}$ . The six cones  $G_4^{ij} = \{\mathbf{h} \in \Gamma_4 | \text{Ingleton}_{ij} \leq 0\}$  are symmetric due to the permutation of inequalities  $\text{Ingleton}_{ij}$ , so it is sufficient to study only one of the cones. Furthermore, [26] gave the extreme rays of  $G_4^{ij}$  in Lemma 1 by using  $\mathbf{r}_t^{\mathcal{K}}$ ,  $\mathbf{g}_i^{(2)}$ ,  $\mathbf{g}_i^{(3)}$  and the following functions  $\mathbf{f}_{ij}$ :

For  $\mathcal{N} = \{1, 2, 3, 4\}$ , define  $\mathbf{f}_{ij}(\mathcal{W})$  as follows:

$$\mathbf{f}_{ij}(\mathcal{W}) = \begin{cases} 3 & \text{if } \mathcal{W} \in \{ik, jk, il, jl, kl\} \\ \min\{4, 2|\mathcal{W}|\} & \text{otherwise} \end{cases}$$

**Lemma 1:** (Matúš) [26] The cone  $G_4^{ij} = \{\mathbf{h} \in \Gamma_4 | \text{Ingleton}_{ij} \leq 0, i, j \in \mathcal{N} \text{ distinct}\}$  is the convex hull of 15 extreme rays. They are generated by the 15 linearly independent functions  $\mathbf{f}_{ij}$ ,  $\mathbf{r}_1^{ijk}$ ,  $\mathbf{r}_1^{ijl}$ ,  $\mathbf{r}_1^{ikl}$ ,  $\mathbf{r}_1^{jkl}$ ,  $\mathbf{r}_1^\emptyset$ ,  $\mathbf{r}_3^\emptyset$ ,  $\mathbf{r}_1^i$ ,  $\mathbf{r}_1^j$ ,  $\mathbf{r}_1^{ik}$ ,  $\mathbf{r}_1^{jk}$ ,  $\mathbf{r}_1^{il}$ ,  $\mathbf{r}_1^{jl}$ ,  $\mathbf{r}_2^k$ ,  $\mathbf{r}_2^l$ , where  $kl = \mathcal{N} \setminus ij$ .

Note that among the 15 extreme rays of  $G_4^{ij}$ , 14 extreme rays  $\mathbf{r}_1^{ijk}$ ,  $\mathbf{r}_1^{ijl}$ ,  $\mathbf{r}_1^{ikl}$ ,  $\mathbf{r}_1^{jkl}$ ,  $\mathbf{r}_1^\emptyset$ ,  $\mathbf{r}_3^\emptyset$ ,  $\mathbf{r}_1^i$ ,  $\mathbf{r}_1^j$ ,  $\mathbf{r}_1^{ik}$ ,  $\mathbf{r}_1^{jk}$ ,  $\mathbf{r}_1^{il}$ ,  $\mathbf{r}_1^{jl}$ ,  $\mathbf{r}_2^k$ ,  $\mathbf{r}_2^l$  are also extreme rays of  $\mathcal{I}_4$  and thus entropic, which leaves  $\mathbf{f}_{ij}$  the only extreme ray in  $G_4^{ij}$  that is not entropic [26]. It is easily verified that  $\bar{\Gamma}_4^*$  is known as long as we know the structure of six cones  $\bar{\Gamma}_4^* \cap G_4^{ij}$ . Due to symmetry, we only need to focus on one of the six cones  $\bar{\Gamma}_4^* \cap G_4^{34}$ , thus we define  $P_4^{34} = \bar{\Gamma}_4^* \cap G_4^{34}$ . Thus, the remainder of the paper aims, in part, to study the properties of supports and probability distributions yielding entropic vectors in  $P_4^{34}$ , this gap between the best known inner and outer bounds for region of entropic vectors on four variables.

Next, let's examine the relationship between subset of the extreme rays of  $G_4^{34}$  and some particular  $p$ -representable semimatroids. There are 15 extreme rays in  $G_4^{34}$ :  $\mathbf{r}_1^{13}$ ,  $\mathbf{r}_1^{23}$ ,  $\mathbf{r}_1^{123}$ ,  $\mathbf{r}_1^{124}$ ,  $\mathbf{r}_1^{134}$ ,  $\mathbf{r}_1^{234}$ ,  $\mathbf{r}_1^\emptyset$ ,  $\mathbf{r}_1^3$ ,  $\mathbf{r}_1^4$ ,  $\mathbf{r}_1^{14}$ ,  $\mathbf{r}_1^2$ ,  $\mathbf{r}_1^{24}$ ,  $\mathbf{r}_2^3$ ,  $\mathbf{f}_{34}$ . We verified that none of the 56  $(i, j | \mathcal{K})$  couples is satisfied by all of the 15 extreme rays. If we remove  $\mathbf{r}_1^{24}$ , then  $(1, 3|2)$  is the only relation satisfied by all the rest 14 extreme rays; if we remove both  $\mathbf{r}_1^{24}$  and  $\mathbf{r}_1^{14}$ , then two relations  $\{(1, 3|2) \ \& \ (2, 3|1)\}$  are satisfied by all the rest 13 extreme rays; at last if we remove  $\mathbf{r}_1^{24}$ ,  $\mathbf{r}_1^{14}$  and  $\mathbf{r}_1^3$ , then the remaining 12 extreme rays all satisfy the set of three relations  $\{(1, 3|2) \ \& \ (2, 3|1) \ \& \ (1, 2|3)\}$ . From Theorem 1, relations  $\{(1, 2|3)\}$ ,  $\{(1, 2|3) \ \& \ (2, 3|1)\}$  and  $\{(1, 2|3) \ \& \ (1, 3|2) \ \& \ (2, 3|1)\}$  as semimatroids



are all *p-representable*, we can say that the faces of  $\Gamma_4$  generated by the corresponding subset of the extreme rays all have interior points that are *p-representable*. Furthermore, as we will see in V-B, the probability distributions associated with these faces of  $\Gamma_4$  are well characterized in Information Geometry.

### III. LISTING CANONICAL $k$ -ATOM SUPPORTS

A first key question when studying the part of the entropy region associated with the gap between its best known inner and outer bounds described in the prior section is which supports for joint probability mass functions for the random variables can yield entropic vectors in this region. In fact, results from Chan [31] have shown that, with infinite computational power, to determine the whole entropy region, it would suffice to consider the conic hull of entropic vectors associated with only the *quasi-uniform* probability distributions, which are completely specified via their support. This question is further motivated by the observation that previous campaigns that have attempted to numerically map unknown parts of this region have empirically observed that the joint probability mass functions associated with the extremal entropic vectors produced, while not quasi-uniform, do have many of their probabilities zero [16], [32].

To begin our study in this arena, we must formally introduce the concept of a  $k$ -atom support and define the equivalence of two  $k$ -atom supports. Consider the probability distributions for a random vector  $\mathbf{X} = (X_1, \dots, X_N)$  taking values on the Cartesian product  $\mathcal{X}^\times = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_N$ , where  $\mathcal{X}_n$  is a finite set with values  $i \in \{1, \dots, |\mathcal{X}_n|\}$ . To a particular probability mass function  $p_{\mathbf{X}}$  we can associate a length  $\prod_{n=1}^N |\mathcal{X}_n| - 1$  vector by listing the probabilities of all but one of the outcomes in  $\mathcal{X}^\times$  into a vector

$$\boldsymbol{\eta} = \left[ \begin{array}{c|c} p_{\mathbf{X}}(i_1, \dots, i_N) & \begin{array}{l} i_k \in \{1, 2, \dots, |\mathcal{X}_k|\}, \\ k \in \{1, \dots, N\}, \\ \sum_{k=1}^N i_k \neq N. \end{array} \end{array} \right]. \quad (15)$$

Notice we are only listing outcomes such that  $\sum_k i_k \neq N$ ,  $p_{\mathbf{X}}(\mathbf{1}) = p_{\mathbf{X}}(i_1 = 1, \dots, i_N = 1)$  will be the only outcome that is left out.  $\boldsymbol{\eta}$  in (15) can be determined uniquely from the probability mass function  $p_{\mathbf{X}}$ , and owing to the fact that the probability mass function must sum to one, the omitted probability  $p_{\mathbf{X}}(\mathbf{1})$  can be calculated, and hence the probability mass function can be determined from  $\boldsymbol{\eta}$ .

Specifying the joint probability distribution via the vector  $\boldsymbol{\eta}$  enables all outcomes to have nonzero probabilities, however, entropic vectors are often extremized by selecting some of the elements of  $\mathcal{X}^\times$  to have zero probability. For this reason, rather than specifying some probabilities in their cartesian product to be zero, it is of interest to instead specify a support  $\mathcal{X} \subset \mathcal{X}^\times$ , no longer a cartesian product, on which the probabilities will be non-zero. Equivalently, if we take  $|\mathcal{X}| = k$ , we are considering only those probability spaces  $(\Omega, \mathcal{F}, \mathbb{P})$  with  $|\Omega| = k$  to define the random variables  $\mathbf{X} : \Omega \rightarrow \mathcal{X}^\times$  on. A probability support  $\mathcal{X}$  satisfying  $|\mathcal{X}| = k$  is called a  $k$ -atom support, and a joint distribution created this way will be called a  $k$ -atom distribution.

Two  $k$ -atom supports  $\mathcal{X}, \mathcal{X}'$ ,  $|\mathcal{X}| = |\mathcal{X}'| = k$ , will be said to be *equivalent*, for the purposes of tracing out the entropy region, if they yield the same set of entropic vectors, up to a permutation of the random variables. In

other words,  $\mathcal{X}$  and  $\mathcal{X}'$  are equivalent, if, for every probability mass function  $p_{\mathbf{X}} : \mathcal{X} \rightarrow [0, 1]$ , there is another probability mass function  $p_{\mathbf{X}'} : \mathcal{X}' \rightarrow [0, 1]$  and a bijection  $\pi : \mathcal{N} \rightarrow \mathcal{N}$  such that

$$h_{\mathcal{A}}(p_{\mathbf{X}}) = h_{\pi(\mathcal{A})}(p_{\mathbf{X}'}) \quad \forall \mathcal{A} \subseteq \mathcal{N}.$$

Take  $N = 4$  and  $|\mathcal{X}| = 1$  as a trivial example, since we only have one outcome/atom, it will have the probability of 1. In this way, different 1-atom supports like  $[(0, 0, 0, 0)]$ ,  $[(0, 0, 1, 2)]$ ,  $[(0, 1, 2, 3)]$  and  $[(2, 5, 7, 9)]$  are equivalent because they all map to the same 15 dimensional entropic vector with all zero elements.

The goal of this section is to formalize this notion of equivalent supports with the use of tools from abstract algebra, then describe some methods for enumerating and listing one representative from each equivalence class of supports.

In this regard, the following definitions from the theory of group actions will be helpful.

**Definition 2:** Let  $Z$  be a finite group acting on a finite set  $\mathcal{V}$ , a *group action* is a mapping

$$Z \times \mathcal{V} \rightarrow \mathcal{V} : (z, v) \mapsto zv$$

such that if  $e$  is the identity in  $Z$ ,  $ev = v \forall v \in \mathcal{V}$  and for any  $z_1, z_2 \in Z$ ,  $z_2 z_1 v = (z_2 z_1) v$  for all  $v \in \mathcal{V}$ . For  $v \in \mathcal{V}$ , the *orbit* of  $v$  under  $Z$  is defined as

$$Z(v) = \{zv \mid z \in Z\}$$

the *stabilizer* subgroup of  $v$  in  $Z$  is defined as

$$Z_v = \{z \in Z \mid zv = v\}$$

Suppose there is some ordering of  $\mathcal{V}$ , and let  $v$  be the element of  $Z(v)$  that is least under this ordering, i.e. the *canonical representative* of the orbit  $Z(v)$ . For another  $v' \in Z(v)$ , an element  $z \in Z$  is called a *transporter element* for  $v'$  if  $zv' = v$ .

**Definition 3:** (orbit data structure [33]) Let  $Z$  be a group which acts on the finite set  $\mathcal{V}$ . The triple

$$\text{orbit}(Z, \mathcal{V}) = (\mathcal{T}, \sigma, \varphi)$$

is the *orbit data structure* for  $Z$  acting on  $\mathcal{V}$  provided that

1.  $\mathcal{T}$  is a transversal of the  $Z$ -orbits on  $\mathcal{V}$
2.  $\sigma : \mathcal{V} \rightarrow L(Z) : v \mapsto Z_v$
3.  $\varphi : \mathcal{V} \rightarrow Z : v \mapsto z$  with  $zv \in \mathcal{T}$

Here,  $L(Z)$  denotes the lattice of subgroups of  $Z$ , we call  $\sigma$  the *stabilizer map* and  $\varphi$  the *transporter map*.

In next section, we will show that listing non-isomorphic distribution supports is equivalent to calculating the orbit data structure associated with the symmetric group acting a particular finite set.

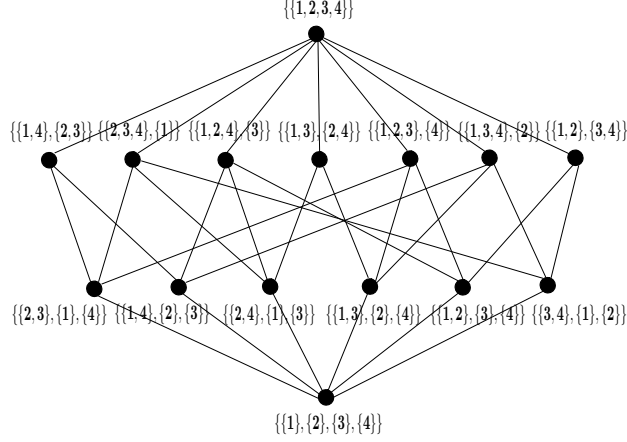


Fig. 1. Lattice of  $\Pi(\mathbb{N}_1^4)$ : the set of all set partitions for  $k = 4$

#### A. Non-isomorphic $k$ -atom supports via Snakes and Ladders

The key to list non-isomorphic distribution supports is to realize that a random variable on a probability space with  $|\Omega| = k$  can be viewed as a *set partition* [34] of  $\mathbb{N}_1^k = \{1, \dots, k\}$  for the purpose of calculating entropy. A *set partition* of  $\mathbb{N}_1^k$  is a set  $\mathcal{B} = \{B_1, \dots, B_t\}$  consisting of  $t$  subsets  $B_1, \dots, B_t$  of  $\mathbb{N}_1^k$ , called the *blocks* of the partition, that are pairwise disjoint  $B_i \cap B_j = \emptyset, \forall i \neq j$ , and whose union is  $\mathbb{N}_1^k$ , so that  $\mathbb{N}_1^k = \bigcup_{i=1}^t B_i$ . Let  $\Pi(\mathbb{N}_1^k)$  denote the set of all set partitions of  $\mathbb{N}_1^k$ . The cardinality of  $\Pi(\mathbb{N}_1^k)$  is commonly known as *Bell numbers*. For instance, there are 5 different set partitions for  $k = 3$ , that is  $|\Pi(\mathbb{N}_1^3)| = 5$  and

$$\begin{aligned} \Pi(\mathbb{N}_1^3) = \{ & \{\{1, 2, 3\}\}, \{\{1, 2\}, \{3\}\}, \{\{1, 3\}, \{2\}\}, \\ & \{\{2, 3\}, \{1\}\}, \{\{1\}, \{2\}, \{3\}\} \}, \end{aligned}$$

while for  $k = 4$ ,  $|\Pi(\mathbb{N}_1^4)| = 15$  and  $\Pi(\mathbb{N}_1^4)$  is the set

$$\begin{aligned} & \{\{\{1, 2, 3, 4\}\}, \{\{1, 2, 3\}, \{4\}\}, \{\{1, 2, 4\}, \{3\}\}, \\ & \{\{1, 3, 4\}, \{2\}\}, \{\{2, 3, 4\}, \{1\}\}, \{\{1, 2\}, \{3, 4\}\}, \\ & \{\{1, 3\}, \{2, 4\}\}, \{\{1, 4\}, \{2, 3\}\}, \{\{1, 2\}, \{3\}, \{4\}\}, \\ & \{\{1, 3\}, \{2\}, \{4\}\}, \{\{1, 4\}, \{2\}, \{3\}\}, \{\{2, 3\}, \{1\}, \{4\}\}, \\ & \{\{2, 4\}, \{1\}, \{3\}\}, \{\{3, 4\}, \{1\}, \{2\}\}, \{\{1\}, \{2\}, \{3\}, \{4\}\} \}. \end{aligned}$$

A set partition  $\mathcal{B} \in \Pi(\mathbb{N}_1^k)$  is said to *refine* a set partition  $\mathcal{B}' \in \Pi(\mathbb{N}_1^k)$  if all of the blocks in  $\mathcal{B}'$  can be written as the union of some blocks in  $\mathcal{B}$ . The *meet* of two partitions  $\mathcal{B}, \mathcal{B}' \in \Pi(\mathbb{N}_1^k)$ , denoted by  $\mathcal{B} \wedge \mathcal{B}'$  is the partition of  $\mathbb{N}_1^k$  formed by all of the non-empty intersections of a block from  $\mathcal{B}$  and a block from  $\mathcal{B}'$ :

$$\mathcal{B} \wedge \mathcal{B}' = \{B_i \cap B'_j \mid B_i \in \mathcal{B}, B'_j \in \mathcal{B}', B_i \cap B'_j \neq \emptyset\}$$

Refinement and meet set up a partial order on  $\Pi(\mathbb{N}_1^k)$  which enable it to be identified as a *lattice*, the lattice of  $\Pi(\mathbb{N}_1^4)$  is shown in Figure 1

Let  $\Xi_N$  be the collection of all sets of  $N$  set partitions of  $\mathbb{N}_1^k$  whose meet is the finest partition (the set of singletons),

$$\Xi_N := \left\{ \xi \left| \xi \subseteq \Pi(\mathbb{N}_1^k), |\xi| = N, \bigwedge_{\mathcal{B} \in \xi} \mathcal{B} = \bigcup_{i=1}^N \{\{i\}\} \right. \right\}. \quad (16)$$

The symmetric group  $\mathbb{S}_k$  induces a natural group action on a set partition  $\mathcal{B} \in \Pi(\mathbb{N}_1^k)$ ,  $\mathcal{B} = \{B_1, \dots, B_t\}$ : representing an element  $\pi \in \mathbb{S}_k$  as a permutation  $\pi : \mathbb{N}_1^k \rightarrow \mathbb{N}_1^k$ , we have

$$\pi(\mathcal{B}) := \{\pi(B_1), \dots, \pi(B_t)\}. \quad (17)$$

This action on set partitions induces, again in a natural manner, a group action of  $\mathbb{S}_k$  on the set  $\Xi_N$  of subsets of  $N$  partitions from  $\Pi(\mathbb{N}_1^k)$  whose meet is the singletons:  $\pi \in \mathbb{S}_k$  acts on a set of partitions  $\xi \in \Xi_N$ ,  $\xi = \{\mathcal{B}_1, \dots, \mathcal{B}_N\}$  via

$$\pi(\xi) := \{\pi(\mathcal{B}_1), \dots, \pi(\mathcal{B}_N)\}. \quad (18)$$

The group action (17) on set partition and group action (18) on sets of set partitions enable us to enumerate the non-isomorphic  $k$ -atom supports by calculating the orbit data structure of symmetry group acting on a well defined set, the result is summarized in Theorem 2.

**Theorem 2:** The problem of generating the list of all non-isomorphic  $k$ -atom,  $N$ -variable supports, that is, selecting one representative from each equivalence class of isomorphic supports, is equivalent to obtaining a transversal of the orbits  $\Xi_N // \mathbb{S}_k$  of  $\mathbb{S}_k$  acting on  $\Xi_N$ , the set of all subsets of  $N$  set partitions of the set  $\mathbb{N}_1^k$  whose meets are the set of singletons  $\{\{1\}, \{2\}, \dots, \{N\}\}$ .

**Proof:** A random variable introduces a partition on the sample space based on its inverse image. The joint distributions of several random variables is created from the meet of these partitions. The joint entropy is insensitive to the labeling of elements in the sample space, as well as the labeling of the outcomes of the random variable, hence it is only the partitions and their meets that matter in determining the joint entropies. Since an appropriate notion of isomorphism between supports also recognizes that it does not matter which random variable is labeled as the first random variable and so on, and there is no need to duplicate random variables when enumerating supports, rather than a  $N$  tuple of partitions, the support of a collection of random variables is best thought of, then, as a set of such set-partitions. The requirement that the meet is the singletons follows from the fact that if it is not, there is a  $k'$  atom distribution with  $k' < k$  whose atoms are the sets in the meet partition, which gives equivalent entropies, and hence such a situation is better labelled as a  $k'$ -atom distribution. ■

Theorem 2 sets up the theoretical framework for obtaining the list of non-isomorphic  $k$ -atom,  $N$ -variable supports for the purpose of calculating entropic vectors: one must calculate the orbits data structure of the symmetric group  $\mathbb{S}_k$  acting on  $\Xi_N$ . One way to carry out this computation is to directly calculate the orbit data structure on  $\Xi_N$  using the default subroutine in GAP. However, this approach quickly becomes intractable when  $k$  and  $N$  are larger than four, as both CPU time and memory usage go beyond the reasonable capacity of a single computer. Alternatively,

$N \setminus k$	3	4	5	6	7
2	2	8	18	48	112
3	2	31	256	2437	25148
4	1	75	2665	105726	5107735
5	0	132	22422	3903832	
6	0	187	161118		

TABLE I  
# OF NON-ISOMORPHIC  $k$ -ATOM,  $N$ -VARIABLE SUPPORTS.

one can use a recursive breadth-first search style algorithm named Leiterspiel or “Snakes and Ladders” [33], [35] to efficiently calculate the orbit data structure.

Suppose we have a group  $Z$  acting on a set  $\mathcal{V}$ , the algorithm Snakes and Ladders, see e.g. [33] pp. 709–710, is an algorithm which enables one to compute orbit data structure of group  $Z$  on the set  $\mathcal{P}_i(\mathcal{V})$  of all subsets of the set  $\mathcal{V}$  of cardinality  $i$ . For a given set  $\mathcal{V}$ , the algorithm first computes the orbit data structure on the set of subsets of size  $i = 1$ , then it recursively increase the subsets size  $i$ , where the computation to determine the orbit data structure for subsets of size  $i$  is created from manipulations with the orbit data structure on subsets of size  $i - 1$ .

To apply this problem to the non-isomorphic support enumeration problem, one selects the set  $\mathcal{V}$  to be the set of all set partitions of the set  $\mathbb{N}_1^k$ , ordered lexicographically, and the group  $Z$  to be the symmetric group  $\mathbb{S}_k$ . Once the step in Snakes and Ladders associated with subsets (of set partitions) of size  $N$  is reached, each element of the transversal is checked to determine if the meet of its partitions is the set of singletons, and the resulting sets of canonical  $N$  set partitions yield the non-isomorphic supports.

The snakes and ladders algorithm was applied to enumerate non-isomorphic supports in this manner, and the resulting numbers of non-isomorphic supports obtained are displayed in Table I. As shown, for  $N = 4$  variables, only one support is needed for calculating entropic vectors of 3-atom distribution, however, there are 75 non-isomorphic supports for  $k = 4$ , and the number of non-isomorphic supports grow rapidly in the number of atoms  $k$ .

In the next section, we will utilize these non-isomorphic  $k$ -atom supports together with numerical optimization to obtain inner bounds for entropy.

#### IV. MAPPING THE ENTROPY REGION WITH $k$ -ATOM DISTRIBUTIONS

With the list of canonical  $k$  supports obtained through the method described in the previous section in hand, the matter turns to how to exploit them to better numerically map out the unknown parts of the entropy region. In this section, we study this problem from two angles, the first, in §IV-A, aims to solely focus on optimizing the Ingleton score, while the second, in §IV-B describes a process for obtaining numerically optimized inner bounds to the entropy region.

number of atoms $k$	3	4	5	6	7
all supports	1	75	2665	105726	5107735
Ingleton violating	0	1	29	1255	60996

TABLE II

# OF NON-ISOMORPHIC  $k$ -ATOM, 4-VARIABLE SUPPORTS THAT CAN VIOLATE THE INGLETON INEQUALITY.

#### A. Maximal Ingleton Violation and the Four Atom Conjecture

Given that the unknown part of  $\bar{\Gamma}_4^*$  is associated with violating the Ingleton inequality, substantial research effort has been exerted towards determining distributions on  $N = 4$  random variables that violate the Ingleton inequality (11). Dougherty, Freiling, and Zeger [32] defined a normalized function called the *Ingleton score* to measure the degree of Ingleton violation for 4 random variables, and they also make the *Four-Atom Conjecture* which states that the Ingleton score of 4 random variables can not be lower than  $-0.08937$ . After the Four-Atom conjecture was proposed, Ingleton violation was studied extensively with finite groups [36]–[38], then in [16], the conjecture was refuted by transforming a distribution obtaining Ingleton score of  $-0.078277$  through a operation which preserves the property of almost entropic to a vector with Ingleton score  $-0.09243$ . In this section, we study the number of  $k$  atom supports that can, for some probability distribution, violate Ingleton, as well as the Ingleton scores they can attain.

For a particular  $k$  atom support for  $N$  variables, the Ingleton score can be numerically optimized via fine grid search and numerical gradient optimization. Doing so for each four variable support with 7 or fewer atoms yielded the results in Table II, which shows that only a small fraction of the canonical supports can violate Ingleton.

Among all the 75 non-isomorphic 4-atom distribution supports, only one can be assigned a series of probabilities to violate Ingleton, that is the 4-atom support (19),

$$\begin{bmatrix} (0, 0, 0, 0) \\ (0, 1, 1, 0) \\ (1, 0, 1, 0) \\ (1, 1, 1, 1) \end{bmatrix}, \quad (19)$$

which is the support achieving Ingleton score  $-0.08937$  associated with the Four atom conjecture.

Each row in (19) is a vector in  $\mathcal{X}$  and corresponds to one outcome/atom for the collection of random variables, so the number of columns in (19) is the same as the number of random variables, in this case 4. Among the 29 5-atom supports that can violate Ingleton, 28 of them obtain the same minimal Ingleton score of  $-0.08937$ , with one atom's probability shrinking to zero. These 28 thus all shrink to the 4-atom support (19), achieving the same

Ingleton score. The remaining one support,

$$\begin{bmatrix} (0, 0, 0, 0) \\ (0, 0, 1, 1) \\ (0, 1, 1, 0) \\ (1, 0, 1, 0) \\ (1, 1, 1, 0) \end{bmatrix} \quad (20)$$

only achieves a minimal Ingleton score of  $-0.02423$ . For the 1255 6-atom supports, 58 of them get a minimal Ingleton score strictly less than  $-0.08937$ , while the remainder of the supports yield minimal scores of  $-0.08937$ . Experiments with the 7-atom supports have shown that, in keeping with the findings of the four atom conjecture and the attempts to refute it, no  $k$ -atom support with  $k \leq 7$  is capable of directly beat the four atom distributions score. These exhaustive results substantiate findings from other researchers that suggest that if it is indeed possible to give a probability distribution which directly (without any almost entropic preserving transformation on the entropic vectors as utilized in [16]) violates the four atom conjecture, at least a large support will be required.

#### B. Optimizing Inner Bounds to Entropy from $k$ -Atom Distributions

For the purpose of generating better inner bounds for the region of entropic vectors, minimizing only the Ingleton score is far from enough, since it is only optimizing the distribution to a cost function of certain hyperplane defined by the Ingleton inequality. Bearing this in mind, one can define cost functions different from the Ingleton score, but still yielding optimized points that are in the unknown part of the entropy region associated with violating Ingleton. We will first describe a simple procedure to randomly generate such cost functions, and then their numerical optimization over each of the Ingleton violating 4,5, and 6 atom supports. The resulting entropic vectors are then collected to generate inner bounds to  $\bar{\Gamma}_4^*$  based on distributions with 4,5, and 6 atom supports.

Lemma 1 defined the 15 extreme rays of the pyramid  $G_4^{ij}$ , and, without loss of generality, it suffices to consider  $G_4^{34}$ . Among these 15 rays, the 14 extreme rays that lie on the hyperplane of  $Ingleton_{34} = 0$  are  $\mathbf{r}_1^{134}, \mathbf{r}_1^{234}, \mathbf{r}_1^{123}, \mathbf{r}_1^{124}, \mathbf{r}_1^\emptyset, \mathbf{r}_3^\emptyset, \mathbf{r}_1^3, \mathbf{r}_1^4, \mathbf{r}_1^{13}, \mathbf{r}_1^{14}, \mathbf{r}_1^{23}, \mathbf{r}_1^{24}, \mathbf{r}_2^1, \mathbf{r}_2^2$ , while the only extreme ray in  $G_4^{34}$  that is not entropic is  $\mathbf{f}_{34}$ . For generating cost functions, among several options, we found the following one gives us the best inner bound. First a random vector  $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_{14}\}$  of 14 dimension is generated, where  $\lambda_i$  takes value from 0 to a large positive value. Then for each of the 14 extreme rays on the hyperplane of  $Ingleton_{34} = 0$ , one new ray is generated through the following equation

$$\mathbf{r}_i^{new} = \frac{\mathbf{r}_i^{base} + \lambda_i \mathbf{f}_{34}}{1 + \lambda_i} \quad (21)$$

where  $\mathbf{r}_i^{base} \in \{\mathbf{r}_1^{134}, \mathbf{r}_1^{234}, \mathbf{r}_1^{123}, \mathbf{r}_1^{124}, \mathbf{r}_1^\emptyset, \mathbf{r}_3^\emptyset, \mathbf{r}_1^3, \mathbf{r}_1^4, \mathbf{r}_1^{13}, \mathbf{r}_1^{14}, \mathbf{r}_1^{23}, \mathbf{r}_1^{24}, \mathbf{r}_2^1, \mathbf{r}_2^2\}$ . After obtaining the 14 new rays  $\mathbf{r}^{new} = \{\mathbf{r}_1^{new}, \dots, \mathbf{r}_{14}^{new}\}$ , the hyperplane defined by these new rays, which in turn defines a new cost function, can be easily calculated. Notice if we let  $\lambda_i = 0$  for  $i = 1, 2, \dots, 14$ , we will get the hyperplane of  $Ingleton_{34} = 0$ .

Computer experiments were run to generate more than 1000 cost functions in this manner by random selection of the  $\lambda_i$ s. For each of these cost functions, numerical optimization of the distribution for each Ingleton violating

inner and outer bounds	percent of pyramid
Shannon	100
Outer bound from [27]	96.5
4,5,6 atoms inner bound	57.8
4,5 atoms inner bound	57.1
4 atoms inner bound	55.9
4 atom conjecture point only	43.5
3 atoms inner bound	0

TABLE III  
THE VOLUME INCREASE WITHIN PYRAMID  $G_4^{34}$  AS MORE ATOMS ARE INCLUDED

$k$ -atom support was performed, and a  $k$ -atom inner bound was generated by taking the convex hull of the entropic vectors corresponding to the optimized distribution.

The progress of characterizing  $\Gamma_4^*$  while performing these experiments was in part estimated by the volume of the inner bound as compared to the total volume of the pyramid  $G_4^{34}$ , as summarized in Table III. For the purpose of comparison, we also list there the calculated volume ratio of the best outer bound in [32]. Note the volume of the  $k$ -atom inner bound obtained through the process describe above is only a estimated value and a lower bound to the true volume fraction, because only a finite number of cost functions and a finite number of entropic vectors were generated through the random cost function generation process. In principle one can generate as many entropic vectors as one wants through this process by growing the number of random cost functions selected, however calculating volume for many extreme points in high dimension can become computationally intractable. A key observation from the process is that while growing the support helps, from a volume standpoint the improvement after four atoms is somewhat small.

However, volume is just one metric for an inner bound, which can also be hard to visualize in high dimensions. For this reason, we would like to visualize the  $k$ -atom inner bound in lower dimension. In this regard, a certain 3 dimensional subset of it selected in [16] will be utilized. In order to perform the transformation, each numerically obtained 15 dimensional vector  $\mathbf{h} \in G_4^{34}$  is first transformed into its *tight* component by subtracting its *modular* component which is defined by

$$\mathbf{h}^m(\mathcal{W}) = \sum_{i \in I} [\mathbf{h}(N) - \mathbf{h}(N \setminus i)] \quad \mathcal{W} \subseteq \mathcal{N}$$

Next  $\mathbf{h}^{ti}$  was pushed onto the hyperplane such that  $I(X_3, X_4) = 0$  and  $I(X_1; X_2 | X_3, X_4) = 0$  through the linear mapping

$$\begin{aligned} \mathbf{h}_{AB} &= A_{34} B_{34,1} \mathbf{h}^{ti} \\ &= \mathbf{h}^{ti} + (h_3^{ti} + h_4^{ti} - h_{34}^{ti})(\mathbf{r}_1^3 - \mathbf{r}_1^\emptyset) \\ &\quad + (h_{123}^{ti} + h_{124}^{ti} - h_{34}^{ti} - h_{1234}^{ti})(\mathbf{r}_2^1 - \mathbf{r}_3^\emptyset) \end{aligned}$$



After that, another linear mapping  $C_{34}$  is used to further reduce the dimension of  $G_4^{34}$  to 4.

$$\begin{aligned}
\mathbf{h}_C &= C_{34}\mathbf{h}_{AB} = -\text{Ingleton}_{34}(\mathbf{h}^{ti})\mathbf{f}_{34} \\
&+ (h_3^{ti} + h_4^{ti} - h_{34}^{ti})\mathbf{r}_1^\emptyset + (h_{123}^{ti} + h_{124}^{ti} - h_{34}^{ti} - h_{1234}^{ti})\mathbf{r}_3^\emptyset \\
&+ \frac{1}{2}(h_{13}^{ti} + h_{23}^{ti} - h_3^{ti} - h_{123}^{ti} + h_{14}^{ti} + h_{24}^{ti} - h_4^{ti} - h_{124}^{ti})(\mathbf{r}_1^3 + \mathbf{r}_1^4) \\
&+ \frac{1}{2}(h_{13}^{ti} + h_{14}^{ti} - h_1^{ti} - h_{134}^{ti} + h_{23}^{ti} + h_{24}^{ti} - h_2^{ti} - h_{234}^{ti})(\mathbf{r}_2^1 + \mathbf{r}_2^2) \\
&+ \frac{1}{4}(h_{12}^{ti} + h_{14}^{ti} - h_1^{ti} - h_{124}^{ti} + h_{12}^{ti} + h_{13}^{ti} - h_1^{ti} - h_{123}^{ti} + \\
&h_{12}^{ti} + h_{24}^{ti} - h_2^{ti} - h_{124}^{ti} + h_{12}^{ti} + h_{23}^{ti} - h_2^{ti} - h_{123}^{ti})(\mathbf{r}_1^{13} + \\
&\mathbf{r}_1^{14} + \mathbf{r}_1^{23} + \mathbf{r}_1^{24})
\end{aligned}$$

If we further normalize the last dimension to equal to one, by dividing through by it, the resulting polytope associated with the convex hull in the remaining three dimensions is three dimensional. In order to coordinatize it, define  $\alpha = \frac{1}{4}\mathbf{f}_{34}$ ,  $\beta = \frac{1}{2}(\mathbf{r}_1^3 + \mathbf{r}_1^4)$ ,  $\gamma = \frac{1}{4}(\mathbf{r}_2^1 + \mathbf{r}_2^2)$  and  $\delta = \frac{1}{4}(\mathbf{r}_1^{13} + \mathbf{r}_1^{14} + \mathbf{r}_1^{23} + \mathbf{r}_1^{24})$ , for any given transformed  $\mathbf{g}$ , we can write

$$\mathbf{g} = \bar{\alpha}_h\alpha + \bar{\beta}_h\beta + \bar{\gamma}_h\gamma + \bar{\delta}_h\delta$$

where  $\bar{\alpha}_h + \bar{\beta}_h + \bar{\gamma}_h + \bar{\delta}_h = 1$ . So in three dimensional space, we consider  $\alpha$  to be  $(0, 0, 0)$ ,  $\beta$  to be  $(\frac{1}{2}, \frac{\sqrt{3}}{2}, 0)$ ,  $\gamma$  to be  $(1, 0, 0)$ ,  $\delta$  to be  $(\frac{1}{2}, \frac{\sqrt{3}}{6}, \frac{\sqrt{6}}{3})$ , so we can make the plot using  $\bar{\beta}_h + \bar{\delta}_h$ ,  $\bar{\gamma}_h + \bar{\delta}_h$  and  $\bar{\alpha}_h$  as the three coordinate.

See Figure 2 for a surface plot of the inner bound generated by 4-atom supports, where we also plot the extremal points of 5-atom (red X) and 6-atom (black squares) inner bounds for comparison. Since we are transforming entropic vector from 15 dimension to 3 dimension, lots of extreme points of our 15 dimensional inner bound actually become redundant in this three dimensional space, so the number of points we can plot is significantly less than the number of extreme points we get from numerical optimization. As can be seen in Figure 2, the extreme points of 4-atom inner bound mostly lies in a curve, and there are some 5-atom extreme points away from the convex hull generated with this curve, and some of 6-atom extreme points can get even further away.

In order to better visualize the difference between the various inner bounds, we also compared the contour of inner bound generated by  $\leq k$  atom supports for  $k \in \{4, 5, 6\}$ , see Figure 3 for this comparison plot where *blue* line is  $k = 4$ , *red* line is  $k = 5$  and *black* line is  $k = 6$ . As you can see, the contour is larger as more atoms are involved, meaning we can constantly get better inner bounds by increasing the number of atoms. This is because as the number of atoms are increased, a greater and greater variety of distribution supports are found that can violate the Ingleton inequality. The increase of the inner bound from five to six atom is smaller than the increase from four to five atoms, which is consistent with the full dimensional volume calculation from Table III.

## V. PERSPECTIVES ON HOW TO USE INFORMATION GEOMETRY TO UNDERSTAND ENTROPY GEOMETRY

The previous section utilized support enumeration followed by numerical optimization to map the unknown parts of the entropy. While this did provide some numerical insight into which supports can have distributions that violate

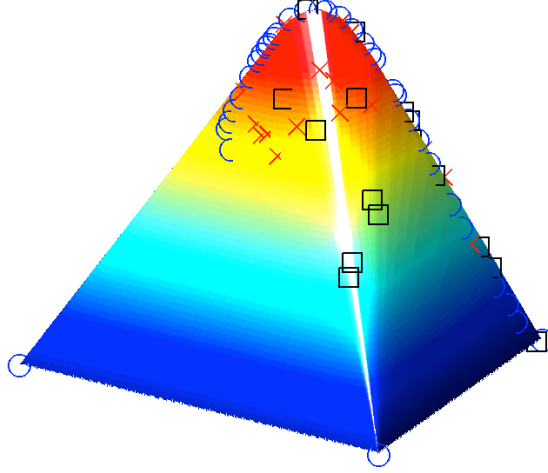


Fig. 2. The surface is a projection of a 3D face of the inner bound to  $\bar{\Gamma}_4^*$  created with the numerical process described in the text with 4-atom distributions. The blue circles are the from the entropic vectors from the extremal optimized four atom distributions, while the red  $X$ 's and the black squares are the additional extremal optimized  $k$  distributions for  $k \in \{5, 6\}$ , respectively.

Ingleton and map to the unknown part of the entropy region, it did not give any analytical insight into how to parametrize probabilities so that the associated entropic vector would be in this unknown part, or be extremal. Bearing this in mind, in this section we aim to study properties of probability distributions associated with entropic vectors that are extremal, in the sense that they give entropic vectors lying in faces of the Shannon outer bound, and also that violate Ingleton. To do so, we will make use of information geometry, a discipline that in part studies ways to coordinatize probability distributions in a manner that relates properties such as independence to affine subspaces of the parameter space.

Information geometry, is a discipline in statistics which endows the manifold of probability distribution with a special “dually flat” differential geometric structure created by selecting a Riemannian metric based on the Fisher information and a family of affine connections called the  $\alpha$ -connections.

The natural divergence functions arising in information geometry, which include the Kullback Leibler divergence, can easily be related to the Shannon entropy by measuring the divergence with the uniform distribution. Additionally, one could think of entropic vectors as characterizing simultaneous properties (entropies) of marginal distributions

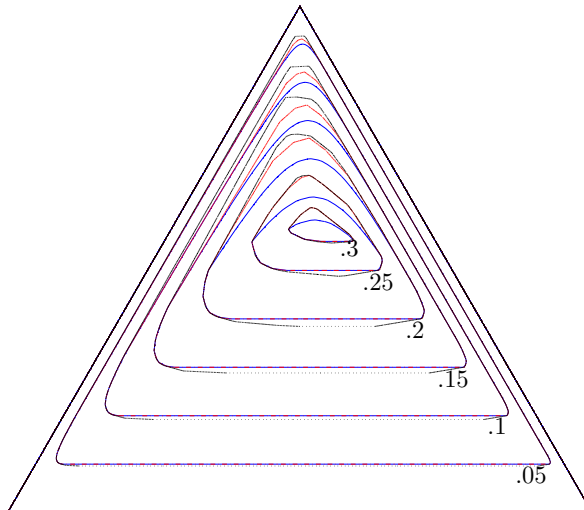


Fig. 3. Comparison of contour plots of inner bound created from  $\leq k$  atom distributions for  $k \in \{4, 5, 6\}$ . For each contour value, the inner most line is  $k = 4$  while the outermost line is  $k = 6$ . The numerical inner bounds generated from only four atom distributions are quite good in this space.

on subsets of variables, and this simultaneous marginal distribution relationship has a well studied structure in information geometry.

Building on these observations, here we wish to sketch out a couple of ideas in an attempt to show that characterizing the boundary of  $\bar{\Gamma}_N^*$  could be thought of as a problem in information geometry, and that information geometry may in fact be a convenient framework to utilize to calculate these boundaries. The perspective we will provide is by no means unique, however, the aim here is to provide a preliminary link between these disciplines and enough evidence about their relationship in the hope that we or other researchers in these areas may one day solve the apparently very difficult, yet fundamental and very important, problem of explicitly determining  $\bar{\Gamma}_N^*$  by exploiting information geometric tools.

Along these lines, we will first review in a somewhat rough manner in §V-A some concepts from information geometry related to entropic vectors. Next, we will provide in §V-B an information geometric characterization of those probability distributions associated with Shannon facets and some Shannon faces of the region of entropic vectors. We then present an information geometric characterization of submodularity as a corollary. Finally, we will close the paper in §V-C with characterization of the 4 atom distributions (19) which violate Ingleton via Information Geometry, and show a natural information geometric parameterization for them.

#### A. Review of Some Relevant Ideas from Information Geometry

Information geometry endows a manifold of probability distributions  $p(x; \xi)$ , parameterized by a vector of real numbers  $\xi = [\xi_i]$ , with a Riemannian metric, or inner product between tangent vectors, given by the Fisher

information:

$$g_{i,j}(\xi) \triangleq \mathbb{E}_{\xi} \left[ \frac{\partial \log p(x; \xi)}{\partial \xi_i} \frac{\partial \log p(x; \xi)}{\partial \xi_j} \right] \quad (22)$$

This allows us to calculate an inner product between two tangent vectors  $c = \sum_i c_i \partial_{\xi_i}$  and  $d = \sum_i d_i \partial_{\xi_i}$  at a particular point  $\xi$  in the manifold of probability distributions, as

$$\langle c, d \rangle_{\xi} = \sum_{i,j} c_i d_j g_{i,j}(\xi) \quad (23)$$

Selecting this Riemannian metric (indeed, we could have selected others), and some additional structure, allows the differential geometric structure to be related to familiar properties of exponential families and mixture families of probability distributions. The additional structure, called an affine connection, is given to the manifold to allow us establish a correspondence, called parallel translation, between tangent vectors living the tangent space at two different points along a curve in the manifold by solving a system of ordinary differential equations involving given functions on the manifold called connection coefficients. Just as with a Riemannian metric, we have an infinite number possible of choices for the affine connection (embodied by selecting connection coefficients with respect to a particular parametrization), but if we choose particular ones, called the  $\alpha$ -connections, certain differential geometric notions like flatness and autoparallel submanifolds can be related to familiar notions of probability distribution families/subfamilies like exponential families and mixture families. While the resulting theory is very elegant, it is also somewhat complex, and hence we must omit the general details, referring the interested reader to [39], and only introduce a small subset of the concepts that can be utilized via a series of examples.

In particular, let's focus our attention on the manifold  $\mathcal{O}(\mathcal{X})$  of probability distributions for a random vector  $\mathbf{X} = (X_1, \dots, X_N)$  taking values on the Cartesian product  $\mathcal{X}^{\times} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_N$ . We already defined a  $\prod_{n=1}^N |\mathcal{X}_n| - 1$  dimensional vector in (15) as  $\boldsymbol{\eta}$ -coordinates, which we also call  $m$ -coordinate in information geometry. Alternatively we could parameterize the probability mass function for such a joint distribution with a vector  $\boldsymbol{\theta}$ , whose  $\prod_{n=1}^N |\mathcal{X}_n| - 1$  elements take the form

$$\boldsymbol{\theta} = \left[ \log \left( \frac{p_{\mathbf{X}}(i_1, \dots, i_N)}{p_{\mathbf{X}}(1, \dots, 1)} \right) \middle| \begin{array}{l} i_k \in \{1, 2, \dots, |\mathcal{X}_k|\}, \\ k \in \{1, \dots, N\}, \\ \sum_{k=1}^N i_k \neq N. \end{array} \right] \quad (24)$$

where  $\mathcal{X}_n$  is a finite set with values denoted by  $i_k$ . These coordinates provide an alternate unique way of specifying the joint probability mass function  $p_{\mathbf{X}}$ , called the  $e$ -coordinates or  $\boldsymbol{\theta}$  coordinates.

A subfamily of these probability mass functions associated with those  $\boldsymbol{\eta}$  coordinates that take the form

$$\boldsymbol{\eta} = \mathbf{A}\mathbf{p} + \mathbf{b} \quad (25)$$

for some  $\mathbf{p}$  for any particular fixed  $\mathbf{A}$  and  $\mathbf{b}$ , that is, that lie in an affine submanifold of the  $\boldsymbol{\eta}$  coordinates, are said to form a  $m$ -autoparallel submanifold of probability mass functions. This is not a definition, but rather a consequence of a theorem involving a great deal of additional structure which must omit here [39].

Similarly, a subfamily of these probability mass functions associated with those  $\boldsymbol{\theta}$  coordinates that take the form

$$\boldsymbol{\theta} = \mathbf{A}\boldsymbol{\lambda} + \mathbf{b} \quad (26)$$

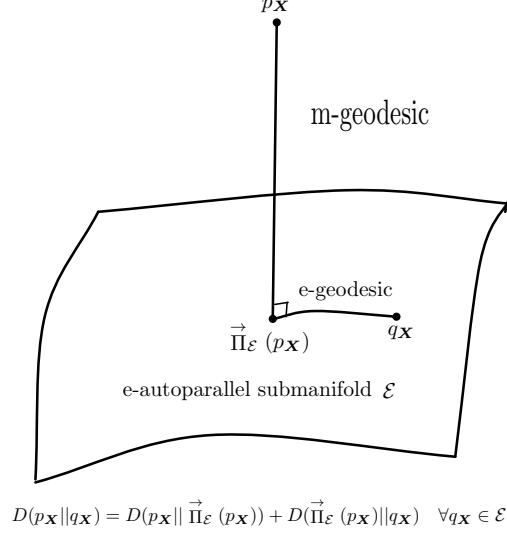


Fig. 4. Pythagorean style relation.

for some  $\lambda$  for any particular fixed  $\mathbf{A}$  and  $\mathbf{b}$ , that is, that lie in an affine submanifold of the  $\theta$  coordinates, are said to form a *e-autoparallel* submanifold of probability mass functions.

An e-autoparallel submanifold (resp. m-autoparallel submanifold) that is one dimensional, in that its  $\lambda$  (resp.  $p$ ) parameter vector is in fact a scalar, is called a *e-geodesic* (resp. *m-geodesic*).

On this manifold of probability mass functions for random variables taking values in the set  $\mathcal{X}$ , we can also define the Kullback Leibler divergence, or relative entropy, measured in bits, according to

$$D(p_{\mathbf{X}}||q_{\mathbf{X}}) = \sum_{\mathbf{x} \in \mathcal{X}} p_{\mathbf{X}}(\mathbf{x}) \log_2 \left( \frac{p_{\mathbf{X}}(\mathbf{x})}{q_{\mathbf{X}}(\mathbf{x})} \right) \quad (27)$$

Note that in this context  $D(p||q) \geq 0$  with equality if and only if  $p = q$ , and hence this function is a bit like a distance, however it does *not* in general satisfy symmetry or triangle inequality.

Let  $\mathcal{E}$  be a particular e-autoparallel submanifold, and consider a probability distribution  $p_{\mathbf{X}}$  not necessarily in this submanifold. The problem of finding the point  $\vec{\Pi}_{\mathcal{E}}(p_{\mathbf{X}})$  in  $\mathcal{E}$  closest in Kullback Leibler divergence to  $p_{\mathbf{X}}$  defined by

$$\vec{\Pi}_{\mathcal{E}}(p_{\mathbf{X}}) \triangleq \arg \min_{q_{\mathbf{X}} \in \mathcal{E}} D(p_{\mathbf{X}}||q_{\mathbf{X}}) \quad (28)$$

is well posed, and is characterized in the following two ways ( here  $\vec{\Pi}_{\mathcal{E}}$  with a right arrow means we are minimizing over the second argument  $q_{\mathbf{X}}$  ). The tangent vector of the m-geodesic connecting  $p_{\mathbf{X}}$  to  $\vec{\Pi}_{\mathcal{E}}(p_{\mathbf{X}})$  is orthogonal, in the sense of achieving Riemannian metric value 0, at  $\vec{\Pi}_{\mathcal{E}}(p_{\mathbf{X}})$  to the tangent vector of the e-geodesic connecting  $\vec{\Pi}_{\mathcal{E}}(p_{\mathbf{X}})$  and any other point  $q \in \mathcal{E}$ . Additionally, for any other point  $q \in \mathcal{E}$ , we have the Pythagorean like relation

$$D(p_{\mathbf{X}}||q_{\mathbf{X}}) = D(p_{\mathbf{X}}||\vec{\Pi}_{\mathcal{E}}(p_{\mathbf{X}})) + D(\vec{\Pi}_{\mathcal{E}}(p_{\mathbf{X}})||q_{\mathbf{X}}). \quad (29)$$

This relationship, which is an important one in information geometry [39], is depicted in Fig. 4.

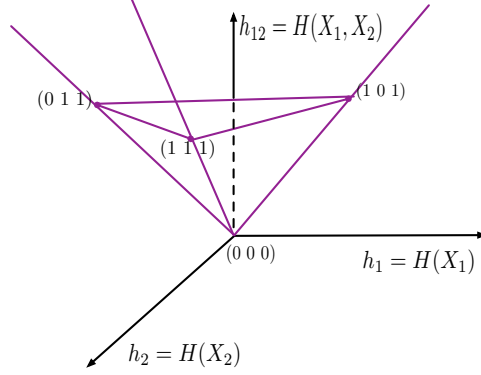


Fig. 5. Region of entropic vector  $\Gamma_2^*$ .

### B. Information Geometric Structure of the Shannon Faces of the Region of Entropic Vectors

As identified in §II-B,  $\Gamma_2 = \Gamma_2^*$  and  $\Gamma_3 = \bar{\Gamma}_3^*$ , implying that  $\Gamma_2^*$  and  $\bar{\Gamma}_3^*$  are fully characterized by Shannon type information inequalities. For example, when  $N = 2$ ,  $\Gamma_2^*$  is a 3-dimensional polyhedral cone characterized by  $H(X_1, X_2) - H(X_1) \geq 0$ ,  $H(X_1, X_2) - H(X_2) \geq 0$  and  $I(X_1; X_2) = H(X_1) + H(X_2) - H(X_1, X_2) \geq 0$  as depicted in Fig. 5.

For  $N = 4$ , even though  $\Gamma_4 \neq \bar{\Gamma}_4^*$  and the region is not a polyhedral cone, there are still many exposed faces of  $\bar{\Gamma}_N^*$  defined by attaining equality in a particular Shannon type information inequality of the form (3) or (4). Such exposed faces of  $\bar{\Gamma}_N^*$  could be referred to as the “Shannon facets” of entropy, and in this section we will first aim to characterize the distributions associated with these Shannon facets via information geometry.

Let  $\mathcal{E}_{\mathcal{A}}^\perp$  be the submanifold of probability distributions for which  $\mathbf{X}_{\mathcal{A}}$  and  $\mathbf{X}_{\mathcal{A}^c} = \mathbf{X}_{\mathcal{N} \setminus \mathcal{A}}$  are independent

$$\mathcal{E}_{\mathcal{A}}^\perp = \{p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}_{\mathcal{A}}}(\mathbf{x}_{\mathcal{A}})p_{\mathbf{X}_{\mathcal{A}^c}}(\mathbf{x}_{\mathcal{A}^c}) \quad \forall \mathbf{x}\} \quad (30)$$

then we define  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp$ ,  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$  and  $\mathcal{M}_{\mathcal{A}}$  as follows

$$\begin{aligned} \mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp &= \{p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = p_{\mathbf{X}_{(\mathcal{A} \cup \mathcal{B})}} p_{\mathbf{X}_{(\mathcal{A} \cup \mathcal{B})^c}}\} \\ \mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp} &= \left\{ p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = p_{\mathbf{X}_{\mathcal{A}}} p_{\mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} \mid \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}} p_{\mathbf{X}_{(\mathcal{A} \cup \mathcal{B})^c}} \right\} \\ \mathcal{M}_{\mathcal{A}}^w &= \{p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = p_{\mathbf{X}_{\mathcal{A}}} \cdot \delta_{\mathcal{A}^c}^w \} \end{aligned}$$

where  $\delta_{\mathcal{A}^c}^w = \begin{cases} 1 & \text{if } \mathbf{X}_{\mathcal{A}^c} = w(\mathbf{X}_{\mathcal{A}}) \\ 0 & \text{otherwise} \end{cases}$  for some fixed function  $w : \mathcal{X}_{\mathcal{A}} \rightarrow \mathcal{X}_{\mathcal{A}^c}$ . Note  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$  is a submanifold of  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp$  such that the random variables  $\mathbf{X}_{\mathcal{A} \cup \mathcal{B}}$ , in addition to being independent from  $\mathbf{X}_{(\mathcal{A} \cup \mathcal{B})^c}$  form the Markov chain  $\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}} \leftrightarrow \mathbf{X}_{\mathcal{A} \cap \mathcal{B}} \leftrightarrow \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}}$ . These sets of distributions will be useful because  $I(\mathbf{X}_{\mathcal{A} \cup \mathcal{B}}; \mathbf{X}_{(\mathcal{A} \cup \mathcal{B})^c}) = h_{\mathcal{A} \cup \mathcal{B}} + h_{(\mathcal{A} \cup \mathcal{B})^c} - h_{\mathcal{N}} = 0$  for every distribution in  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp$  and  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$ ,  $I(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} \mid \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) = h_{\mathcal{A}} + h_{\mathcal{B}} -$

$h_{(\mathcal{A} \cap \mathcal{B})} - h_{(\mathcal{A} \cup \mathcal{B})} = 0$  for every distribution in  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$ , and  $h_{\mathcal{N}} - h_{\mathcal{A}} = 0$  for every distribution in  $\mathcal{M}_{\mathcal{A}}^w$  for any  $w$ .

We will show that the sets of distributions  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^{\perp}$ ,  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$ ,  $\mathcal{M}_{\mathcal{A}}^w$  are of interest because they correspond to every Shannon facet. This is because every facet of  $\Gamma_N$ , which must correspond to equality in an inequality of the form (3) or (4), can be regarded as some conditional entropy, mutual information, or conditional mutual information being identically zero. For example, as depicted in Fig. 5,  $\Gamma_2^*$  has three facets corresponding to set of distributions  $\mathcal{E}_{1 \cup 2}^{\perp}$ ,  $\{\mathcal{M}_1^w\}$  and  $\{\mathcal{M}_2^w\}$  respectively. Surprisingly, while the entropy vector is a nonlinear function of the joint distribution, and these families of distributions correspond to the intersection of affine sets with the region of entropic vectors, they are themselves, when regarded with the correct information geometric parameterization, *associated with affine sets of distributions*.

In order to prove the information geometric properties of Shannon facets, we first give the equivalent condition of  $I(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) = 0$  in distribution space, especially in  $\boldsymbol{\theta}$  coordinate as mentioned in (26).

**Lemma 2:** Let  $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$ ,  $\mathcal{A}, \mathcal{B} \subset \mathcal{N} = \{1, 2, \dots, N\}$  where  $|\mathcal{X}_{\mathcal{A} \setminus \mathcal{B}}| = m$ ,  $|\mathcal{X}_{\mathcal{B} \setminus \mathcal{A}}| = n$  and  $|\mathcal{X}_{\mathcal{A} \cap \mathcal{B}}| = q$ . For each assignment  $C_0$  of  $\mathbf{X}_{\mathcal{A} \cap \mathcal{B}}$ , define the following equations for  $i = 1, 2, \dots, m-1$  and  $j = 1, 2, \dots, n-1$  for the probabilities  $p_{\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}}, \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}}, \mathcal{X}_{\mathcal{A} \cap \mathcal{B}}}$ :

$$\frac{p_{0jC_0} p_{i0C_0}}{p_{00C_0} p_{ijC_0}} = 1 \quad (31)$$

which is equivalent with hyperplane (32) in  $\boldsymbol{\theta}$  coordinates for  $p_{\mathbf{X}_{\mathcal{A} \cup \mathcal{B}}}$ ,

$$\theta_{0jC_0} + \theta_{i0C_0} - 1[C_0 \neq \mathbf{0}] \cdot \theta_{00C_0} - \theta_{ijC_0} = 0 \quad (32)$$

then  $I(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) = 0$  and the set of equations (31), (32) are equivalent.

**Proof:** From Equation (31), it can be verified that

$$\frac{p_{wjC_0} p_{ieC_0}}{p_{weC_0} p_{ijC_0}} = 1 \quad (33)$$

for any distinct  $i, j, w$  and  $e$  such that  $w, i \in \{0, 1, \dots, m-1\}$  and  $j, e \in \{0, 1, \dots, n-1\}$ . More specifically, the equations in (33) but not in (31) all can be derived by combining some of the  $(m-1)(n-1)$  equations in (31).

$\Rightarrow$  Since  $I(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) = 0$ , for each assignment  $C_0$  of  $\mathbf{X}_{\mathcal{A} \cap \mathcal{B}}$ , it can be verified [1] that

$$\begin{aligned} & p_{\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}} \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}} (x_{\mathcal{A} \setminus \mathcal{B}} = i, x_{\mathcal{B} \setminus \mathcal{A}} = j, x_{\mathcal{A} \cap \mathcal{B}} = C_0) \\ &= p_{\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}} (x_{\mathcal{A} \setminus \mathcal{B}} = i | x_{\mathcal{A} \cap \mathcal{B}} = C_0) \\ & p_{\mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}} (x_{\mathcal{B} \setminus \mathcal{A}} = j | x_{\mathcal{A} \cap \mathcal{B}} = C_0) \end{aligned}$$

which make sure the nominator and denominator cancelled with each other in (31) and (33).

$\Leftarrow$  Now suppose (31), (32) and (33) hold for  $w, i \in \{0, 1, \dots, m-1\}$  and  $j, e \in \{0, 1, \dots, n-1\}$ , it suffices to show for each assignment  $C_0$  of  $\mathbf{X}_{\mathcal{A} \cap \mathcal{B}}$ ,

$$H(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) + H(\mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) = H(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}}, \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) \quad (34)$$

We can calculate:

$$H(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}} = C_0) = \sum_{i=0}^{m-1} [(\sum_{j=0}^{n-1} p_{ijC_0}) \log(\sum_{j=0}^{n-1} p_{ijC_0})] \quad (35)$$

$$H(\mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}} = C_0) = \sum_{j=0}^{n-1} [(\sum_{i=0}^{m-1} p_{ijC_0}) \log(\sum_{i=0}^{m-1} p_{ijC_0})] \quad (36)$$

$$H(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}}, \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}} = C_0) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} p_{ijC_0} \log p_{ijC_0} \quad (37)$$

From (35) and (36) we have

$$\begin{aligned} H(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}} = C_0) + H(\mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}} = C_0) = \\ \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} p_{ijC_0} \log[(\sum_{e=0}^{n-1} p_{ieC_0})(\sum_{w=0}^{m-1} p_{wjC_0})] \end{aligned} \quad (38)$$

where  $(\sum_{e=0}^{n-1} p_{ieC_0})(\sum_{w=0}^{m-1} p_{wjC_0}) = p_{ijC_0}$  when (33) are satisfied, in which case (34) hold.  $\square$

Now let's take  $I(X_1; X_2 | X_3) = 0$  as a example:

**Example 1:** Let  $\mathbf{X} = \{X_1, X_2, X_3\}$ , where  $|\mathcal{X}_1| = m$ ,  $|\mathcal{X}_2| = n$  and  $|\mathcal{X}_3| = q$ . For  $mnq$  different values of  $\mathbf{X}$ , if we want to parameterize this manifold  $\mathcal{O}_3$  in  $\theta$  coordinate, we will need  $(mnq - 1)$  different parameters:  $\theta_1, \theta_2, \dots, \theta_{mnq-1}$ , which can be defined as in (24). We define submanifold  $\mathcal{E}_3$  to satisfy the constraint  $I(X_1; X_2 | X_3) = 0$ . It is easy to verify that the dimension of  $\mathcal{O}_3$  is  $(mnq - 1)$ , the dimension of  $\mathcal{E}_3$  is  $((m + n - 1)q - 1)$ . We know any hyperplane  $\sum_{i=1}^{mnq-1} \lambda_i \theta_i = 0$  is a  $(mnq - 2)$  dimensional e-autoparallel submanifold in  $\mathcal{O}_3$ . Then subtracting  $((m + n - 1)q - 1) + 1$  from  $mnq - 2$ , we know  $\mathcal{E}_3$  is the intersection of  $(m - 1)(n - 1)q$  different hyperplanes, which are  $\sum_{i=1}^{mnq-1} \lambda_i^k \theta_i = 0$  for  $k = 1, 2, \dots, (m - 1)(n - 1)q$ . For  $i = 1, 2, \dots, m - 1$ ,  $j = 1, 2, \dots, n - 1$  and  $r = 1, 2, \dots, q$ , each hyperplane  $\sum_{i=1}^{mnq-1} \lambda_i^k \theta_i = 0$  corresponding to a constraint on the probability:

$$\frac{p_{0jr} p_{i0r}}{p_{00r} p_{ijr}} = 1 \quad (39)$$

which can be written as

$$\theta_{0jr} + \theta_{i0r} - \mathbb{1}[r \neq 0] \cdot \theta_{00r} - \theta_{ijr} = 0 \quad (40)$$

where  $\mathbb{1}[r \neq 0] = 0$  when  $r = 0$ .

Now we can use Lemma 2 to prove Theorem 3, then use the relationship between  $m$ -autoparallel submanifold and affine subspace to prove Theorem 4:

**Theorem 3:**  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp} \subseteq \mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^{\perp} \subseteq \mathcal{O}(\mathcal{X})$ ,  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$  is an e-autoparallel submanifold of  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^{\perp}$  and  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^{\perp}$  is an e-autoparallel submanifold of  $\mathcal{O}(\mathcal{X})$ .



**Proof:** Follows directly from Bayes' rule, we can rewrite  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp$  and  $\mathcal{O}(\mathcal{X})$  as

$$\begin{aligned}\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp &= \left\{ p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = p_{\mathbf{X}_{\mathcal{A}}} p_{\mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A}}} p_{\mathbf{X}_{(\mathcal{A} \cup \mathcal{B})^c}} \right\} \\ \mathcal{O}(\mathcal{X}) &= \left\{ p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = p_{\mathbf{X}_{(\mathcal{A} \cup \mathcal{B})}} p_{\mathbf{X}_{(\mathcal{A} \cup \mathcal{B})^c} | (\mathcal{A} \cup \mathcal{B})} \right\}\end{aligned}$$

then we can easily verify  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp} \subseteq \mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp \subseteq \mathcal{O}(\mathcal{X})$ . Since from Lemma 2, the equivalent condition of  $I(\mathbf{X}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} | \mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) = 0$  in  $e$ -coordinate for  $p_{\mathbf{X}_{\mathcal{A} \cup \mathcal{B}}}$  is the intersection of a sequences of hyperplanes (32) and the case of  $\mathcal{A} \cap \mathcal{B} = \emptyset$  can be considered as a special case of Lemma 2. Then from the definition of autoparallel submanifold,  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$  is  $e$ -autoparallel in  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp$ , and  $\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp$  is  $e$ -autoparallel in  $\mathcal{O}(\mathcal{X})$ .  $\square$

**Theorem 4:** Let  $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N} = \{1, 2, \dots, N\}$ , then  $\mathcal{M}_{\mathcal{A}}^w \subseteq \mathcal{M}_{\mathcal{B}}^{w'} \subseteq \mathcal{O}(\mathcal{X})$ ,  $\mathcal{M}_{\mathcal{A}}^w$  is a  $m$ -autoparallel submanifold of  $\mathcal{M}_{\mathcal{B}}^{w'}$  and  $\mathcal{M}_{\mathcal{B}}^{w'}$  is a  $m$ -autoparallel submanifold of  $\mathcal{O}(\mathcal{X})$ .

**Proof:** It follows directly from the definition of  $m$ -autoparallel submanifold since for the  $m$ -affine coordinate of  $\mathcal{O}(\mathcal{X})$  and  $\mathcal{M}_{\mathcal{B}}^{w'}$ , we can easily find the corresponding matrix  $A$  and vector  $B$  such that  $\mathcal{M}_{\mathcal{B}}^{w'}$  is affine subspace of  $\mathcal{O}(\mathcal{X})$ ; similarly for the  $m$ -affine coordinate of  $\mathcal{M}_{\mathcal{A}}^w$  and  $\mathcal{M}_{\mathcal{B}}^{w'}$ , we can easily find the corresponding matrix  $\mathbf{A}$  and vector  $\mathbf{b}$  such that  $\mathcal{M}_{\mathcal{A}}^w$  is affine subspace of  $\mathcal{M}_{\mathcal{B}}^{w'}$ .  $\square$

Theorem 3 and Theorem 4, have shown that Shannon facets are associated with affine subsets of the family of probability distribution, when it is regarded in an appropriate parameterization. In fact, as we shall presently show, all Shannon Type information inequalities correspond to the positivity of divergences of projections to these submanifolds. Note that  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$  is set up so that the difference between the right and left sides in the entropy submodularity inequality (3) is zero for every probability distribution in  $\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}$ . The nested nature of these  $e$ -autoparallel submanifolds, and the associated Pythagorean relation, is one way to view the submodularity of entropy, as we shall now explain with Figure 6 and Corollary 2.

**Corollary 2:** The submodularity (3) of the entropy function can be viewed as a consequence of Pythagorean style information projection relationships depicted in Figure 6. In particular, submodularity is equivalent to the inequality

$$D\left(\vec{\Pi}_{\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp}(p_{\mathbf{X}}) \parallel \vec{\Pi}_{\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}}(p_{\mathbf{X}})\right) \geq 0 \quad (41)$$

since

$$\begin{aligned}D\left(\vec{\Pi}_{\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp}(p_{\mathbf{X}}) \parallel \vec{\Pi}_{\mathcal{E}_{\mathcal{A}, \mathcal{B}}^{\leftrightarrow, \perp}}(p_{\mathbf{X}})\right) &= H(\mathbf{X}_{\mathcal{A}}) + H(\mathbf{X}_{\mathcal{B}}) \\ &\quad - H(\mathbf{X}_{\mathcal{A} \cap \mathcal{B}}) - H(\mathbf{X}_{\mathcal{A} \cup \mathcal{B}})\end{aligned}$$

**Proof:** The projections are

$$\vec{\Pi}_{\mathcal{E}_{\mathcal{A} \cup \mathcal{B}}^\perp}(p_{\mathbf{X}}) = p_{\mathbf{X}_{\mathcal{A} \cup \mathcal{B}}} p_{\mathbf{X}_{(\mathcal{A} \cup \mathcal{B})^c}} \quad (42)$$

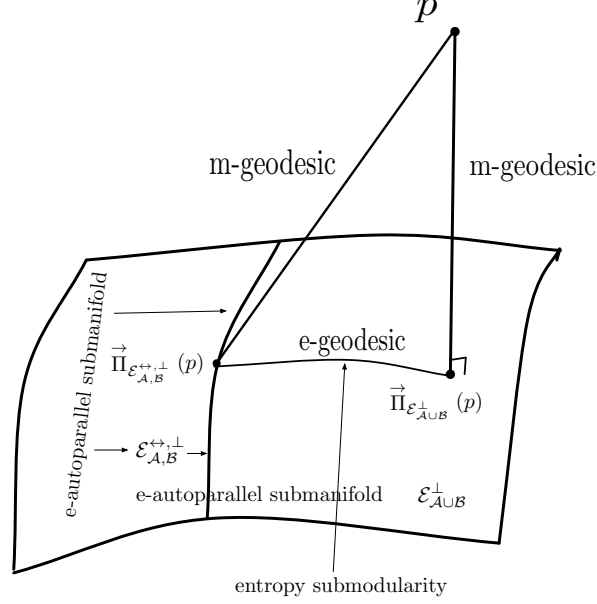


Fig. 6. Submodularity of the entropy function is equivalent to the non-negativity of a divergence  $D(\vec{\pi}_{\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}}(p) || \vec{\pi}_{\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}}(p))$  between two information projections, one projection ( $\vec{\pi}_{\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}}(p)$ ) is to a set that is a submanifold the other projection's ( $\vec{\pi}_{\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}}(p)$ ) set. A Pythagorean style relation shows that for such an arrangement  $D(p || \vec{\pi}_{\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}}(p)) = D(p || \vec{\pi}_{\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}}(p)) + D(\vec{\pi}_{\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}}(p) || \vec{\pi}_{\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}}(p))$ .

and

$$\vec{\Pi}_{\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}}(p_{\mathbf{X}}) = p_{\mathbf{X}_{\mathcal{A}\setminus\mathcal{B}}} | \mathbf{X}_{\mathcal{A}\cap\mathcal{B}} p_{\mathbf{X}_{\mathcal{B}}} p_{\mathbf{X}_{(\mathcal{A}\cup\mathcal{B})^c}} \quad (43)$$

since for  $\vec{\Pi}_{\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}}(p_{\mathbf{X}})$  given by (42) for every  $q_{\mathbf{X}} \in \mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}$  we can reorganize the divergence as

$$\begin{aligned} D(p_{\mathbf{X}} || q_{\mathbf{X}}) &= D(p_{\mathbf{X}} || \vec{\Pi}_{\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}}(p_{\mathbf{X}})) \\ &+ D(\vec{\Pi}_{\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}}(p_{\mathbf{X}}) || q_{\mathbf{X}_{(\mathcal{A}\cup\mathcal{B})}} q_{\mathbf{X}_{(\mathcal{A}\cup\mathcal{B})^c}}) \end{aligned}$$

and for  $\vec{\Pi}_{\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}}(p_{\mathbf{X}})$  given by (43) for every  $q_{\mathbf{X}} \in \mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}$  we can reorganize the divergence as

$$\begin{aligned} D(p_{\mathbf{X}} || q_{\mathbf{X}}) &= D(p_{\mathbf{X}} || \vec{\Pi}_{\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}}(p_{\mathbf{X}})) \\ &+ D(\vec{\Pi}_{\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}}(p_{\mathbf{X}}) || q_{\mathbf{X}_{\mathcal{A}\setminus\mathcal{B}}} | \mathbf{X}_{\mathcal{A}\cap\mathcal{B}} q_{\mathbf{X}_{\mathcal{B}}} q_{\mathbf{X}_{(\mathcal{A}\cup\mathcal{B})^c}}) \end{aligned}$$

The remainder of the corollary is proved by substituting (42) and (43) in the equation for the divergence.  $\square$

In order to further illustrate these ideas, we will demonstrate Theorem 3, Theorem 4 and Corollary 2 in the context of  $\bar{\Gamma}_3^*$ .

**Example 2:** Setting  $N = 3$  we have  $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ . Let  $\mathcal{N} = \{1, 2, 3\}$ , and denote  $\mathcal{A}$  and  $\mathcal{B}$  as some

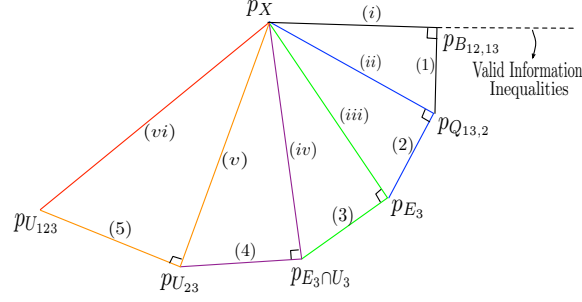


Fig. 7. Pythagorean relationship on 3 discrete random variables

subset of  $\mathcal{N}$  that is not equal to  $\emptyset$  and  $\mathcal{N}$ , then we consider the following submanifolds of  $\mathcal{O}(\mathcal{X})$ :

$$\begin{aligned}
 B_{\mathcal{A},\mathcal{B}} &= \{p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = p_{\mathbf{X}_{\mathcal{A}}} \cdot p_{\mathbf{X}_{\mathcal{B} \setminus \mathcal{A}} \mid (\mathcal{A} \cap \mathcal{B})}\} \\
 Q_{\mathcal{A},\mathcal{B}} &= \{p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = p_{\mathbf{X}_{\mathcal{A}}} \cdot p_{\mathbf{X}_{\mathcal{B}}}\} \\
 E_3 &= \left\{p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = \prod_{i=1}^3 p_{X_i}\right\} \\
 U_{\mathcal{A}} &= \left\{p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}} = \frac{1}{\prod_{i \in \mathcal{A}} |\chi_i|} p_{\mathbf{X}_{\mathcal{A}^c}}\right\}
 \end{aligned}$$

Theorem 3 and Theorem 4 show that  $B_{\mathcal{A},\mathcal{B}}$ ,  $Q_{\mathcal{A},\mathcal{B}}$ ,  $E_3$  and  $U_{\mathcal{A}}$  are all nested e-autoparallel submanifolds of  $\mathcal{O}(\mathcal{X})$ . For example,  $U_{123} \subset U_{23} \subset E_3 \cap U_3 \subset E_3 \subset Q_{13,2} \subset B_{12,13} \subset S_3$ , and in this chain, the previous one is the e-autoparallel submanifold of the later one. If we denote  $p_{B_{12,13}}$ ,  $p_{Q_{13,2}}$ ,  $p_{E_3}$ ,  $p_{E_3 \cap U_3}$ ,  $p_{U_{23}}$  and  $p_{U_{123}}$  the m-projection of a point  $p \in S_3 \setminus B_{12,13}$  onto  $B_{12,13}$ ,  $Q_{13,2}$ ,  $E_3$ ,  $E_3 \cap U_3$ ,  $U_{23}$  and  $U_{123}$  respectively, we will have the Pythagorean relation as shown in Figure 7. From these pythagorean relations, we will get a series of Information inequalities

$$\begin{aligned}
 (i) \quad & h_{12} + h_{13} \geq h_1 + h_{123} \\
 (ii) \quad & h_{13} + h_2 \geq h_{123} \\
 (iii) \quad & h_1 + h_2 + h_3 \geq h_{123} \\
 (iv) \quad & h_1 + h_2 + \log |\chi_3| \geq h_{123} \\
 (v) \quad & h_1 + \log |\chi_2| + \log |\chi_3| \geq h_{123} \\
 (vi) \quad & \log |\chi| \geq h_{123}
 \end{aligned}$$

as well as

- (1)  $h_1 + h_2 \geq h_{12}$
- (2)  $h_1 + h_3 \geq h_{13}$
- (3)  $\log |\chi_3| \geq h_3$
- (4)  $\log |\chi_2| \geq h_2$
- (5)  $\log |\chi_1| \geq h_1$

where each inequality corresponds to the positivity of the divergence with the associated label in Figure 7. As these are the basic Shannon type inequalities for  $N = 3$ , this diagram yields an information geometric interpretation for the region of entropic vectors  $\bar{\Gamma}_3^*$ .

Lemma 2 can also be used to derive information geometric results for certain Shannon faces.

**Theorem 5:** Let relation  $\mathcal{L}' \subseteq \mathcal{S}(N)$  be the containment minimal  $p$ -representable semimatroid containing a set  $\mathcal{L} \subseteq \mathcal{S}(N)$  such that  $\{i \cup j \cup \mathcal{K}\} = \mathcal{N}' \subseteq \mathcal{N}$  for  $\forall (i, j | \mathcal{K}) \in \mathcal{L}$ . Denote  $\mathcal{O}(\mathcal{X}')$  the manifold of probability distributions for random vector  $\mathbf{X}_{\mathcal{N}'}$ . Let  $N' = |\mathcal{N}'|$  and define  $\mathcal{F}_{\mathcal{N}'}$  in (44),

$$\mathcal{F}_{\mathcal{N}'} = \{ \mathbf{h} \in \Gamma_{N'} \mid h_{i\mathcal{K}} + h_{j\mathcal{K}} - h_{\mathcal{K}} - h_{ij\mathcal{K}} = 0, \forall (i, j | \mathcal{K}) \in \mathcal{L}' \} \quad (44)$$

then the set of distributions corresponding to  $\mathcal{F}_{\mathcal{N}'}$  form a e-autoparallel submanifold in  $\mathcal{O}(\mathcal{X}')$ .

**Proof:** From Corollary 1 we know  $\mathcal{F}_{\mathcal{N}'}$  is a face of  $\Gamma_{N'}$ . As a result of Lemma 2, each  $(i, j | \mathcal{K}) \in \mathcal{L}$  correspond to a series of hyperplanes (32) in  $\boldsymbol{\theta}$  coordinate, we use  $E_{eq}^{\mathcal{F}_0}$  to denote the intersection of all the hyperplanes representing any  $(i, j | \mathcal{K})$  pairs in  $\mathcal{L}$ . Since for the submanifold

$$\mathcal{E}^\perp = \left\{ p_{\mathbf{X}}(\cdot) \mid p_{\mathbf{X}}(\mathbf{x}) = \prod_{i \in \mathcal{N}'} p_{X_i}(x_i) \right\}$$

we have  $\mathbf{h}(\mathcal{E}^\perp) \subseteq \mathcal{F}_{\mathcal{N}'}$ , which means the intersection  $E_{eq}^{\mathcal{F}_0}$  will always be non-empty, and hence there will always exist a  $p$ -representable semimatroid containing  $\mathcal{L}$ . The containment minimal  $p$ -representable matroid will contain any other conditional independence relations implied by those in  $\mathcal{L}$ , and hence any point in  $E_{eq}^{\mathcal{F}_0}$  will obey these too. Furthermore,  $E_{eq}^{\mathcal{F}_0}$  is e-autoparallel in  $\mathcal{O}(\mathcal{X}')$  since it is the non-empty intersections of hyperplanes in  $\boldsymbol{\theta}$  coordinate according to Lemma 2.  $\square$

**Example 3:** As a example, recall in the last paragraph of section §II-D we study the relationship between subset of the extreme rays of  $G_4^{34}$  and  $p$ -representable semimatroids  $\{(1, 3|2)\}$ ,  $\{(1, 3|2) \& (2, 3|1)\}$  and  $\{(1, 2|3) \& (1, 3|2) \& (2, 3|1)\}$ . Since all these three semimatroids are  $p$ -representable, there are distributions satisfy the conditional independent relations corresponding to these  $(i, j | \mathcal{K})$  couples. For example, the distributions corresponding to semimatroid  $\{(1, 2|3) \& (1, 3|2) \& (2, 3|1)\}$  satisfied the constraints  $I(X_1; X_2 | X_3) = 0$ ,  $I(X_1; X_3 | X_2) = 0$  and  $I(X_2; X_3 | X_1) = 0$ . Then according to Lemma 2, the distributions

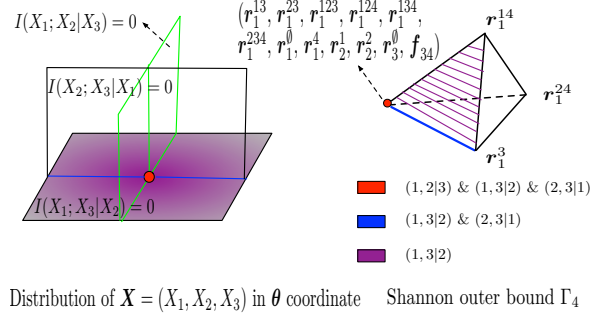


Fig. 8. The mapping between some faces of  $\Gamma_4$  and submanifold of four variable distribution in  $\theta$  coordinate

which corresponding to semimatroid  $\{(1, 2|3) \& (1, 3|2) \& (2, 3|1)\}$  must obey (32) defined for each of the three  $(i, j|\mathcal{K})$  couples. In  $\theta$  coordinate, the mapping can be illustrated by the plot on the left of Fig. 8, in which we only consider three random variables  $X_1, X_2$  and  $X_3$ . Since each of three set of distributions is e-autoparallel, any their intersections are also e-autoparallel. On the right of Fig. 8 we plot the relationship among the Shannon faces corresponding to the three  $p$ -representable semimatroids  $\{(1, 3|2)\}$ ,  $\{(1, 3|2) \& (2, 3|1)\}$  and  $\{(1, 2|3) \& (1, 3|2) \& (2, 3|1)\}$  in the gap  $G_4^{34}$  of  $\Gamma_4$ . As shown in Fig. 8, the polyhedral cone constructed from  $r_1^3, r_1^{14}, r_1^{13}, r_1^{23}, r_1^{123}, r_1^{124}, r_1^{134}, r_1^{234}, r_1^0, r_1^4, r_1^1, r_1^2, r_1^3$  and  $f_{34}$  corresponding to semimatroids  $\{(1, 3|2)\}$ ; removing  $r_1^{14}$  from the list, we get semimatroids  $\{(1, 3|2) \& (2, 3|1)\}$ ; finally, the intersection of three faces corresponding to the polyhedral cone constructed from  $r_1^{13}, r_1^{23}, r_1^{123}, r_1^{124}, r_1^{134}, r_1^{234}, r_1^0, r_1^4, r_1^1, r_1^2, r_1^3$  and  $f_{34}$ .

As for the relations between  $p$ -representable semimatroids and the conditions of  $\mathcal{L}$  specified in Theorem 5, by examine the 120 irreducible  $p$ -representable semimatroids listed in Theorem 1, we get the following Corollary:

**Corollary 3:** Let relation  $\mathcal{L} \subseteq \mathcal{S}(4)$  be a  $p$ -representable semimatroid such that  $\mathcal{L}$  contains at least two  $(i, j|\mathcal{K})$  pairs. If  $\{i \cup j \cup \mathcal{K}\} = \mathcal{N}' \subseteq \{1, 2, 3, 4\}$  for  $\forall (i, j|\mathcal{K}) \in \mathcal{L}$ , then  $\mathcal{L}$  is an Ingleton semimatroid.

In this sense, the information geometric parametrization presented thus far can only reach those faces of  $\bar{\Gamma}_4^*$  shared with the Ingleton inner bound. In part for this reason, in the next section we will study the information geometric structure of certain Ingleton violating distributions.

### C. Information Geometric Structure of Ingleton-Violating Entropic Vectors & their Distributions

As discussed in §IV-A, every  $k$ -atom distribution for  $k < 4$  is incapable of violating Ingleton, and every  $k$ -atom distribution for  $k \in \{5, 6, 7\}$  are incapable of exceeding the Ingleton score achieved by the 4-atom distribution on support  $\mathcal{X}_4$  given in (19). Here we investigate study the information geometric structure of those distributions violating Ingleton on this special 4-atom support, and show that they admit a nice information geometric. We will also show that larger  $k$ -atom supports that are capable of violating Ingleton, but do not yield better Ingleton scores, do not share this special information geometric structure.

Denote by  $\mathcal{O}(\mathcal{X}_4)$  the manifold of all probability mass functions for four binary random variables with the support (19). Parameterize these distributions with the parameters  $\eta_1 = \mathbb{P}(\mathbf{X} = 0000) = \alpha$ ,  $\eta_2 = \mathbb{P}(\mathbf{X} = 0110) = \beta - \alpha$ ,  $\eta_3 = \mathbb{P}(\mathbf{X} = 1010) = \gamma - \alpha$ ,  $\eta_4 = \mathbb{P}(\mathbf{X} = 1111) = 1 - \sum_{i=1}^3 \eta_i = 1 + \alpha - \beta - \gamma$ , yielding the m-coordinates of  $\mathcal{O}(\mathcal{X}_4)$ . Let the associated e-coordinates can be calculated as  $\theta_i = \log_2 \frac{\eta_i}{\eta_4}$  for  $i = 1, 2, 3$ . Next, we consider the submanifold  $\mathcal{D}_{u1} = \{p_x \in \mathcal{O}(\mathcal{X}_4) | I(X_3; X_4) = 0\}$ , by Theorem 3,  $\mathcal{O}(\mathcal{X}_4)$  is a e-autoparallel submanifold of  $\mathcal{D}_{4atom}$ . In fact, an equivalent definition is  $\mathcal{D}_{u1} = \{p_x \in \mathcal{O}(\mathcal{X}_4) | -\theta_1 + \theta_2 + \theta_3 = 0\}$ . Numerical calculation illustrated in the Figure 9 lead to the following proposition, which we have verified numerically.

**Proposition 1:** Let  $\mathcal{D}_{m1} = \{p_x \in \mathcal{O}(\mathcal{X}_4) | Ingleton_{34} = 0\}$ , then  $\mathcal{D}_{m1}$  is a e-autoparallel submanifold of  $\mathcal{O}(\mathcal{X}_4)$  and is parallel with  $\mathcal{D}_{u1}$  in e-coordinate. In fact, an equivalent definition of  $\mathcal{D}_{m1}$  is  $\mathcal{D}_{m1} = \{p_x \in \mathcal{O}(\mathcal{X}_4) | -\theta_1 + \theta_2 + \theta_3 = \log_2(\frac{0.5-\alpha_0}{\alpha_0})^2\}$ , where  $\alpha_0$  is the solution of  $-\alpha \log_2 \alpha - (1 - \alpha) \log_2(1 - \alpha) = \frac{1+2\alpha}{2}$  in  $0 < \alpha < \frac{1}{2}$ .

In fact, using this equivalent definition, we can also determine all the distributions in  $\mathcal{O}(\mathcal{X}_4)$  that violate  $Ingleton_{34} \geq 0$  as the submanifold  $\mathcal{D}_{Vio} = \{p_x \in \mathcal{O}(\mathcal{X}_4) | -\theta_1 + \theta_2 + \theta_3 < \log_2(\frac{0.5-\alpha_0}{\alpha_0})^2\}$ . Because we are dealing with  $\mathcal{O}(\mathcal{X}_4)$ , a 3 dimensional manifold, we can use a plot to visualize our results in Fig. 9. In Fig. 9, besides  $\mathcal{D}_{u1}$  and  $\mathcal{D}_{m1}$ , we also plot the following submanifolds and points:

$$\begin{aligned} \mathcal{D}_{q1} &= \{p_x \in \mathcal{O}(\mathcal{X}_4) | Ingleton_{34} = 0.1\} \\ \mathcal{D}_{s1} &= \{p_x \in \mathcal{O}(\mathcal{X}_4) | Ingleton_{34} = -0.126\} \\ \mathcal{D}_{w1} &= \{p_x \in \mathcal{O}(\mathcal{X}_4) | Ingleton_{34} = -0.16\} \\ \mathcal{D}_{n1} &= \{p_x \in \mathcal{O}(\mathcal{X}_4) | \beta = \gamma = 0.5\} \\ p_u &= \{p_x \in \mathcal{O}(\mathcal{X}_4) | \alpha = 0.25, \beta = \gamma = 0.5\} \\ p_m &= \{p_x \in \mathcal{O}(\mathcal{X}_4) | \alpha \approx 0.33, \beta = \gamma = 0.5\} \end{aligned}$$

As we can see from Fig. 9,  $\mathcal{D}_{m1}$  and  $\mathcal{D}_{u1}$  are e-autoparallel and parallel to each other. As  $Ingleton_{34}$  goes from 0 to negative values, the hyperplane becomes ellipsoid-like, and as  $Ingleton_{34}$  becomes smaller and smaller, the ellipsoid-like surface shrinks, finally shrinking to a single point  $p_m$  at  $Ingleton_{34} \approx -0.1699$ , the point associated with the four atom conjecture in [32]. Also for each e-autoparallel submanifold  $\mathcal{D}_{Ve} \subset \mathcal{D}_{Vio}$  that is parallel to  $\mathcal{D}_{m1}$  in e-coordinate, the minimum Ingleton of  $\mathcal{D}_{Ve}$  is achieved at point  $\mathcal{D}_{Ve} \cap \mathcal{D}_{n1}$ , where  $\mathcal{D}_{n1}$  is the e-geodesic in which marginal distribution of  $X_3$  and  $X_4$  are uniform, i.e.  $\beta = \gamma = 0.5$ .

Next, we map some submanifolds of  $\mathcal{O}(\mathcal{X}_4)$  to the entropic region. From Lemma 1,  $G_4^{34}$ , one of the six gaps between  $\mathcal{J}_4$  and  $\Gamma_4$  is characterized by extreme rays  $\bar{\mathbf{V}}_P = (\bar{\mathbf{V}}_M, \bar{\mathbf{V}}_R, \mathbf{r}_1^0, \mathbf{f}_{34})$ , where  $\bar{\mathbf{V}}_M = (\mathbf{r}_1^{13}, \mathbf{r}_1^{14}, \mathbf{r}_1^{23}, \mathbf{r}_1^{24}, \mathbf{r}_2^1, \mathbf{r}_2^2)$  and  $\bar{\mathbf{V}}_R = (\mathbf{r}_1^{123}, \mathbf{r}_1^{124}, \mathbf{r}_1^{134}, \mathbf{r}_1^{234}, \mathbf{r}_1^3, \mathbf{r}_1^4, \mathbf{r}_3^0)$ . In addition, we define  $\bar{\mathbf{V}}_N = (\mathbf{r}_1^1, \mathbf{r}_1^2, \mathbf{r}_1^{12})$ , and use Fig. 10 to help visualize  $G_4^{34}$ .

In Fig. 10,  $\mathbf{f}_{34}$  is one of the six Ingleton-violating extreme ray of  $\Gamma_4$ ,  $\bar{\mathbf{V}}_M, \bar{\mathbf{V}}_R$  and  $\mathbf{r}_1^0$  are all extreme rays of  $\mathcal{J}_4$  that make  $Ingleton_{34} = 0$ . Based on the information geometric characterization, the mapping from  $\mathcal{O}(\mathcal{X}_4)$  to



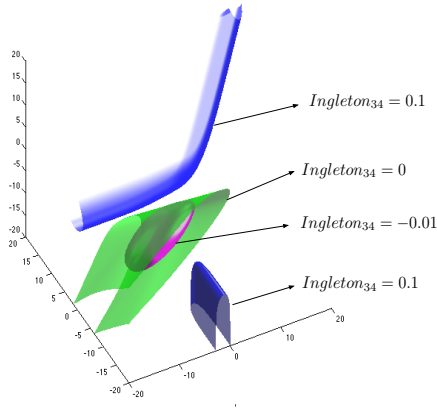


Fig. 11. Slice of  $\mathcal{D}_{5atom}$  in  $\theta$  coordinate

the problem of listing non-isomorphic distribution supports for the purpose of calculating entropic vectors. This is carried out by fixing  $k$ , the number of atoms and  $N$ , the number of random variables, so we can grow in  $k$  or  $N$  to see the progress we make towards the characterization of entropy region. Along the way, a recursive algorithm, Snakes and Ladders, was used to efficiently enumerate the unique supports. The concept of inner bounds based on  $k$ -atom distributions was introduced to aid understanding the structure of the entropic vector region. We experimentally generated  $k$ -atom inner bounds for  $k = 4, 5$ , and  $6$ , calculated the volume of these inner bounds, and visualized them via a certain three dimensional projection. A future research direction in this line of work is to study the conditional independence relations specified by the  $k$ -atom support that violate Ingleton, and to explore the algebraic structure of  $k$ -atom supports.

The second part of the paper shifted away from numerical distribution searches toward analytical characterization of properties of distributions which enabled them to be extremal in the sense achieving entropic vectors of living in certain faces of the Shannon outer bound, as well as for violating Ingleton. These analytical characterizations made use of Information geometric parameterizations. It was shown that the set of distributions on the support associated with the yet-best Ingleton score achievable with an (not-almost) entropic vector which violate Ingleton correspond to a half-space in an appropriate coordinate system. This property was shown not be shared by a larger 5-atom support achieving a strictly poorer Ingleton score.

## REFERENCES

- [1] Raymond W. Yeung, *Information Theory and Network Coding*. Springer, 2008.
- [2] R. W. Yeung and Zhen Zhang, “Distributed source coding for satellite communications,” *IEEE Trans. on Information Theory*, vol. 45, no. 4, pp. 1111–1120, 1999.
- [3] Xijin Yan, Raymond W. Yeung, and Zhen Zhang, “The Capacity Region for Multi-source Multi-sink Network Coding,” in *IEEE International Symposium on Information Theory (ISIT)*, Jun. 2007, pp. 116 – 120.
- [4] —, “An Implicit Characterization of the Achievable Rate Region for Acyclic Multisource Multisink Network Coding,” *IEEE Trans. on Information Theory*, vol. 58, no. 9, pp. 5625–5639, Sep. 2012.



- [5] T. Chan and A. Grant, "Entropy Vectors and Network Codes," in *IEEE International Symposium on Information Theory*, Jun. 2007.
- [6] T. Chan and A. Grant, "Dualities between entropy functions and network codes," in *Fourth Workshop on Network Coding, Theory and Applications (NetCod)*, January 2008.
- [7] —, "Mission impossible: computing the network coding capacity region," in *IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 320–324.
- [8] Zhen Zhang and Raymond W. Yeung, "On Characterization of Entropy Function via Information Inequalities," *IEEE Trans. on Information Theory*, vol. 44, no. 4, Jul. 1998.
- [9] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-Shannon-type inequalities for entropies," *Communication in Information and Systems*, vol. 2, no. 2, pp. 147–166, December 2002.
- [10] R. Dougherty, C. Freiling, and K. Zeger, "Six new non-Shannon information inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, July 2006, pp. 233–236.
- [11] František Matúš, "Infinitely Many Information Inequalities," in *IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2007, pp. 41–44.
- [12] T. Kaced, "Equivalence of two proof techniques for non-shannon-type inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, July 2013.
- [13] B. Hassibi and S. Shadbakht, "On a Construction of Entropic Vectors Using Lattice-Generated Distributions," in *IEEE International Symposium on Information Theory (ISIT)*, Jun. 2007, pp. 501 – 505.
- [14] Weidong Xu, Jia Wang, Jun Sun, "A projection method for derivation of non-Shannon-type information inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, 2008, pp. 2116 – 2120.
- [15] J. M. Walsh and S. Weber, "A Recursive Construction of the Set of Binary Entropy Vectors and Related Inner Bounds for the Entropy Region," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, Oct. 2011.
- [16] F. Matúš and L. Csirmaz, "Entropy region and convolution," Oct. 2013, arXiv:1310.5957v1.
- [17] Yunshu Liu and J. M. Walsh, "Bounding the entropic region via information geometry," in *IEEE Information Theory Workshop*, Seville, Spain, Sep. 2013, pp. 577–581.
- [18] J. M. Walsh and S. Weber, "A Recursive Construction of the Set of Binary Entropy Vectors," in *Forty-Seventh Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2009, pp. 545–552.
- [19] —, "Relationships Among Bounds for the Region of Entropic Vectors in Four Variables," in *2010 Allerton Conference on Communication, Control, and Computing*.
- [20] Congduan Li, John MacLaren Walsh, Steven Weber, "A computational approach for determining rate regions and codes using entropic vector bounds," in *50th Annual Allerton Conference on Communication, Control and Computing*, Oct. 2012. [Online]. Available: <http://www.ece.drexel.edu/walsh/Allerton2012LJW.pdf>
- [21] Babak Hassibi, Sormeh Shadbakht, Matthew Thill, "On Optimal Design of Network Codes," in *Information Theory and Applications*, UCSD, Feb. 2010, presentation.
- [22] Martí-Farré and J. Padró, "On Secret Sharing Schemes, Matroids and Polymatroids," in *4th Theory of Cryptography Conference(TCC)*, Feb. 2007, pp. 273–290.
- [23] L. Csirmaz, "The size of a share must be large," *Journal of Cryptology*, vol. 10, pp. 223–231, Sep. 1997.
- [24] James Oxley, *Matroid Theory*, 2nd. Ed. Oxford University Press, 2011.
- [25] D. Hammer, A. Romashchenko, A. Shen, N. Vereshchagin, "Inequalities for Shannon Entropy and Kolmogorov Complexity," *Journal of Computer and System Sciences*, vol. 60, pp. 442–464, 2000.
- [26] F. Matúš and M. Studený, "Conditional Independences among Four Random Variables I," *Combinatorics, Probability and Computing*, vol. 4, no. 3, pp. 269–278, Sep. 1995.
- [27] Randall Dougherty, Chris Freiling, Kenneth Zeger, "Linear rank inequalities on five or more variables," submitted to SIAM J. Discrete Math. arXiv:0910.0284.
- [28] Ryan Kinser, "New Inequalities for Subspace Arrangements," *J. of Comb. Theory Ser. A*, vol. 188, no. 1, pp. 152–161, Jan. 2011.
- [29] A. W. Ingleton, "Representation of Matroids," in *Combinatorial Mathematics and its Applications*, D. J. A. Welsh, Ed. San Diego: Academic Press, 1971, pp. 149–167.
- [30] F. Matúš, "Conditional Independences among Four Random Variables III: Final Conclusion," *Combinatorics, Probability and Computing*, vol. 8, no. 3, pp. 269–276, May 1999.

- [31] T. Chan and R. Yeung, “On a relation between information inequalities and group theory,” *IEEE Trans. on Information Theory*, vol. 48, no. 7, pp. 1992 – 1995, Jul. 2002.
- [32] Randall Dougherty, Chris Freiling, Kenneth Zeger, “Non-Shannon Information Inequalities in Four Random Variables,” Apr. 2011, arXiv:1104.3602v1.
- [33] A. Betten, M. Braun, H. Friepertinger, A. Kerber, A. Kohnert, and A. Wassermann, *Error-Correcting Linear Codes: Classification by Isometry and Applications*. Springer, 2006.
- [34] D. E. Knuth, *The Art of Computer Programming, Volume 4A*. Addison-Wesley Professional, 2011.
- [35] Bernd Schmalz, “ $t$ -Designs zu vorgegebener Automorphismengruppe,” *Bayreuther Mathematische Schriften*, no. 41, pp. 1–164, 1992, Dissertation, Universität Bayreuth, Bayreuth.
- [36] Wei Mao, Matthew Thill, and Babak Hassibi, “On group network codes: Ingleton-bound violations and independent sources,” in *IEEE International Symposium on Information Theory (ISIT)*, Jun. 2010.
- [37] Nigel Boston and Ting-Ting Nan, “Large Violations of the Ingleton Inequality,” in *50th Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2012.
- [38] Pirita Paajanen, “Finite  $p$ -Groups, Entropy Vectors, and the Ingleton Inequality for Nilpotent Groups,” *IEEE Trans. on Information Theory*, vol. 60, no. 7, pp. 3821–3824, Jul. 2014.
- [39] S. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Society Translations of Mathematical Monographs, 2004, vol. 191.